GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 14:20:36 ; Search time 11.93 Seconds
(without alignments)
1263.404 Million cell updates/sec

Title: US-08-883-036A-2
Perfect score: 440
Sequence: 1 MEQRGQNAPAASGARKRHGP.....HLLSSGKFMYLEGNADSAMS 440

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 6

Total number of hits satisfying chosen parameters: 843

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	1.8	598	1	NU5M_BRALA	O79422 branchiosto
2	8	1.8	599	1	NU5M_BRAFL	O47430 branchiosto
3	8	1.8	1239	1	V120_EBV	P03189 epstein-bar
4	7	1.6	61	1	AKH2_LOCMI	P08379 locusta mig
5	7	1.6	79	1	RL22_PRUAP	Q44161 prunus arme
6	7	1.6	111	1	SINR_BACSU	P06533 bacillus su
7	7	1.6	169	1	GP38_CANFA	Q95152 canis famil
8	7	1.6	174	1	NUC_SHIFL	P29769 shigella fl
9	7	1.6	179	1	RIMM_MYCLE	O33016 mycobacteri
10	7	1.6	197	1	YDEQ_BACSU	P96674 bacillus su
11	7	1.6	243	1	YO27_CAEEL	P34677 caenorhabdi
12	7	1.6	256	1	MTDH_UROFA	O00058 uromyces fa
13	7	1.6	297	1	ARCC_ECOLI	P37306 escherichia
14	7	1.6	308	1	METE_METTH	O26869 methanobact
15	7	1.6	319	1	MER4_STRLI	P30344 streptomyce
16	7	1.6	328	1	HOLB_PSEAE	P52024 pseudomonas
17	7	1.6	333	1	IF2B_RABIT	P41035 oryctolagus
18	7	1.6	354	1	CAHC_HUMAN	O43570 homo sapien
19	7	1.6	361	1	SUR6_HUMAN	O75683 homo sapien
20	7	1.6	379	1	METX_CORGL	O68640 corynebacte
21	7	1.6	381	1	CYCH_RHIME	P45400 rhizobium m
22	7	1.6	395	1	NR23_MOUSE	Q9qxz7 mus musculu
23	7	1.6	397	1	LEF1_MOUSE	P27782 mus musculu
24	7	1.6	397	1	LEF1_RAT	Q9qxn1 rattus norv
25	7	1.6	399	1	LEF1_HUMAN	Q9uju2 homo sapien
26	7	1.6	435	1	Y413_ARATH	O04656 arabidopsis
27	7	1.6	437	1	SYD_METTH	O26328 methanobact
28	7	1.6	522	1	RECN_HELPJ	Q9zj80 helicobacte
29	7	1.6	523	1	EAT3_MOUSE	P51906 mus musculu
30	7	1.6	523	1	EAT3_RAT	P51907 rattus norv
31	7	1.6	524	1	EAT3_HUMAN	P43005 homo sapien
32	7	1.6	539	1	CH61_MYCTU	Q59573 mycobacteri
33	7	1.6	557	1	GSHC_TOBAC	P80461 nicotiana t
34	7	1.6	558	1	ROL_HUMAN	P14866 homo sapien
35	7	1.6	660	1	CCMF_PSEFL	P52225 pseudomonas
36	7	1.6	762	1	P115_CHICK	Q98917 gallus gall
37	7	1.6	794	1	SC18_CANAL	P34732 candida alb
38	7	1.6	956	1	RRPO_SBMV	P21405 southern be
39	7	1.6	1176	1	RPB2_DROME	P08266 drosophila
40	7	1.6	1245	1	POLS_SINDO	P27285 sindbis vir
41	7	1.6	2067	1	BIMB_EMENI	P33144 emericella
42	7	1.6	2204	1	RRPL_NDVB	P11205 newcastle d
43	7	1.6	2326	1	PGG2_RAT	Q00657 rattus norv
44	7	1.6	3054	1	POLG_TEV	P04517 t genome po
45	7	1.6	3255	1	POLG_LMV0	P31999 l genome po
46	7	1.6	3255	1	POLG_LMVE	P89876 l genome po
47	7	1.6	3390	1	POLG_DEN3	P27915 d genome po
48	6	1.4	50	1	VG38_BPMD2	O64229 mycobacteri
49	6	1.4	59	1	HSP1_MACRU	P42142 macropus ru
50	6	1.4	59	1	HSP1_PHACI	P79990 phascolarct
51	6	1.4	60	1	HSP1_NOTTY	P42143 notoryctes
52	6	1.4	63	1	PER_DROMS	Q04535 drosophila
53	6	1.4	65	1	YCF9_CYAPA	P17159 cyanophora
54	6	1.4	71	1	YE35_METJA	Q58830 methanococc
55	6	1.4	74	1	HS12_SOYBN	P02520 glycine max
56	6	1.4	74	1	YIA1_MYCTU	P19771 mycobacteri
57	6	1.4	81	1	YPT2_PSEAE	P24561 pseudomonas
58	6	1.4	82	1	DHBK_SOYBN	O49227 glycine max
59	6	1.4	85	1	IMMC_ECOLI	P02986 escherichia
60	6	1.4	88	1	RS15_THETH	P80378 thermus aqu
61	6	1.4	92	1	PLM_CANFA	P56513 canis famil
62	6	1.4	92	1	PLM_HUMAN	O00168 homo sapien
63	6	1.4	92	1	PLM_RAT	O08589 rattus norv
64	6	1.4	95	1	RS15_STRCO	O86655 streptomyce
65	6	1.4	101	1	LSM3_HUMAN	Q9y4z1 homo sapien
66	6	1.4	102	1	FE45_RHISN	P55542 rhizobium s
67	6	1.4	103	1	EMF1_CHICK	P08317 gallus gall
68	6	1.4	105	1	CRTC_PIG	P28491 sus scrofa
69	6	1.4	105	1	YPT2_STRCO	Q53867 streptomyce
70	6	1.4	106	1	GUAU_RAT	P70668 rattus norv
71	6	1.4	108	1	VMEM_SMYEA	Q00842 strawberry
72	6	1.4	109	1	PER_SYRPI	Q26612 syritta pip
73	6	1.4	109	1	RLA1_MAIZE	P52855 zea mays (m
74	6	1.4	111	1	SINR_BACLI	P22753 bacillus li
75	6	1.4	114	1	RL12_HALHA	P05768 halobacteri
76	6	1.4	115	1	R18E_HALMA	P12733 haloarcula
77	6	1.4	115	1	VMEM_PVX	P07697 potato viru
78	6	1.4	115	1	VMEM_PVXHB	Q07633 potato viru
79	6	1.4	115	1	YT3D_STRCO	Q9xai3 streptomyce
80	6	1.4	117	1	CHH3_PENJP	Q94676 penaeus jap
81	6	1.4	117	1	SP21_BACLI	P26777 bacillus li
82	6	1.4	120	1	RL18_BACHD	Q9z9j8 bacillus ha
83	6	1.4	122	1	PHNG_RHIME	Q52984 rhizobium m
84	6	1.4	122	1	SVS5_MOUSE	P30933 mus musculu
85	6	1.4	122	1	YEBF_ECOLI	P33219 escherichia
86	6	1.4	125	1	TRAC_STAAU	P06698 staphylococ
87	6	1.4	127	1	YQJB_ECOLI	P42615 escherichia
88	6	1.4	128	1	CHEY_SALTY	P06657 salmonella
89	6	1.4	128	1	CYCP_RHOSP	P00146 rhodopseudo
90	6	1.4	128	1	KV3K_HUMAN	P06311 homo sapien
91	6	1.4	129	1	CYCP_RHOCA	P00147 rhodobacter
92	6	1.4	130	1	COTM_BACSU	Q45058 bacillus su
93	6	1.4	134	1	LCTB_BACCA	P06549 bacillus ca
94	6	1.4	134	1	TYRT_STRGA	P55047 streptomyce
95	6	1.4	137	1	YCDP_ECOLI	P75904 escherichia
96	6	1.4	137	1	YQA3_THEAQ	P74897 thermus aqu
97	6	1.4	142	1	NDK_AQUAE	O67528 aquifex aeo
98	6	1.4	143	1	HS11_MEDSA	P27879 medicago sa
99	6	1.4	144	1	RL15_SULAC	O05643 sulfolobus
100	6	1.4	145	1	YMW4_YEAST	Q04773 saccharomyc
101	6	1.4	146	1	TYRT_STRAT	P17687 streptomyce
102	6	1.4	148	1	YORX_PYRWO	P20298 pyrococcus
103	6	1.4	149	1	TRB1_RHISN	P55398 rhizobium s
104	6	1.4	149	1	YHE1_YEAST	P38730 saccharomyc
105	6	1.4	150	1	RS13_ARATH	P49203 arabidopsis
106	6	1.4	151	1	YDHU_BACSU	O05513 bacillus su

107	6	1.4	152	1	RL9_MYCLE	P46385	mycobacteri
108	6	1.4	154	1	HS11_LYCES	P30221	lycopersico
109	6	1.4	154	1	HS15_SOYBN	P04795	glycine max
110	6	1.4	156	1	HS11_ARATH	P19036	arabidopsis
111	6	1.4	158	1	HS12_MEDSA	P27880	medicago sa
112	6	1.4	160	1	TATB_HELPJ	Q9zm58	helicobacte
113	6	1.4	160	1	YGAD_PSEPU	P72227	pseudomonas
114	6	1.4	161	1	HS16_SOYBN	P05478	glycine max
115	6	1.4	162	1	HXA9_CAVPO	P51783	cavia porce
116	6	1.4	162	1	Y096_GVCL	P41728	cryptophleb
117	6	1.4	163	1	YAG5_SCHPO	Q09869	schizosacch
118	6	1.4	164	1	GSPM_ERWCA	P31709	erwinia car
119	6	1.4	166	1	GP38_RAT	Q64294	rattus norv
120	6	1.4	166	1	Y410_SYNY3	Q55116	synechocyst
121	6	1.4	168	1	SSB_BRUAB	Q07432	brucella ab
122	6	1.4	169	1	AROK_CORGL	Q9x5d1	corynebacte
123	6	1.4	172	1	ESP4_LACVV	P35578	lacerta viv
124	6	1.4	172	1	Y074_NPVOP	O10327	orgyia pseu
125	6	1.4	177	1	IMP2_YEAST	P46972	saccharomyc
126	6	1.4	179	1	ATPF_SYNY3	P27181	synechocyst
127	6	1.4	179	1	VG59_HSVEB	P28983	equine herp
128	6	1.4	180	1	PORC_METTH	O27772	methanobact
129	6	1.4	180	1	SMCX_CRIGR	P41228	cricetulus
130	6	1.4	181	1	NDKM_COLLI	P87355	columba liv
131	6	1.4	185	1	GRA2_TOXGO	P13404	toxoplasma
132	6	1.4	185	1	PORC_PYRHO	O58411	p pyruvate/
133	6	1.4	185	1	RRF_STRCO	O86770	streptomyce
134	6	1.4	185	1	THF2_ARATH	Q9xfh9	arabidopsis
135	6	1.4	189	1	RK12_SPIOL	P02398	spinacia ol
136	6	1.4	190	1	ISPZ_BRAJA	P30961	bradyrhizob
137	6	1.4	191	1	COXZ_RHOSH	P56940	rhodobacter
138	6	1.4	192	1	A16_ANOGA	Q93118	anopheles g
139	6	1.4	193	1	MLEC_CHICK	P02606	gallus gall
140	6	1.4	194	1	AX2A_PHAAU	P32293	phaseolus a
141	6	1.4	194	1	KTHY_METTH	O27793	methanobact
142	6	1.4	194	1	MLEV_HUMAN	P08590	homo sapien
143	6	1.4	195	1	ROA1_BOVIN	P09867	bos taurus
144	6	1.4	196	1	LAST_SERMA	P37006	serratia ma
145	6	1.4	198	1	RM12_HUMAN	P52815	homo sapien
146	6	1.4	199	1	MLEV_RAT	P16409	rattus norv
147	6	1.4	201	1	IORB_ARCFU	O28249	archaeoglob
148	6	1.4	201	1	NUC1_RHIME	O68854	rhizobium m
149	6	1.4	203	1	FLA1_ARCFU	O29208	archaeoglob
150	6	1.4	203	1	RM12_CRICR	P52827	cricetus cr
151	6	1.4	204	1	GIDB_COXBU	P94614	coxiella bu
152	6	1.4	206	1	EMBP_CRIGR	P81448	cricetulus
153	6	1.4	207	1	RAB8_HUMAN	P24407	homo sapien
154	6	1.4	208	1	ERA_BRAJA	O69162	bradyrhizob
155	6	1.4	210	1	ENGB_ECOLI	P24253	escherichia
156	6	1.4	210	1	YIHA_ECOLI	P24253	escherichia
157	6	1.4	211	1	FDOI_ECOLI	P32174	escherichia
158	6	1.4	211	1	RS3A_ARCFU	O27964	archaeoglob
159	6	1.4	212	1	HIS5_NEIMA	Q9jvh3	neisseria m
160	6	1.4	212	1	UL45_HSVTH	P18536	turkey herp
161	6	1.4	213	1	CPPB_NEIGO	P07049	neisseria g
162	6	1.4	213	1	EF1B_CANAL	P78590	candida alb
163	6	1.4	213	1	NCCN_ALCXX	Q44587	alcaligenes
164	6	1.4	214	1	N6M1_HUMAN	Q9y5n5	homo sapien
165	6	1.4	215	1	YDQ5_SCHPO	O14198	schizosacch
166	6	1.4	218	1	CTFA_CLOAB	P33752	clostridium
167	6	1.4	218	1	YBGJ_ECOLI	P75744	escherichia
168	6	1.4	219	1	CUT2_MYCTU	Q50664	mycobacteri
169	6	1.4	219	1	YIAD_ECOLI	P37665	escherichia
170	6	1.4	222	1	EMBP_HUMAN	P13727	homo sapien
171	6	1.4	224	1	DKK4_HUMAN	Q9ubt3	homo sapien
172	6	1.4	225	1	CDSN_PIG	O19084	sus scrofa
173	6	1.4	225	1	YB67_AERPE	Q9ycu5	aeropyrum p
174	6	1.4	228	1	COX2_ADABI	P29871	adalia bipu
175	6	1.4	229	1	SNPA_STRSQ	P43163	streptomyce
176	6	1.4	231	1	NU4M_HYPHY	O03733	hypnale hyp
177	6	1.4	233	1	FC3B_HUMAN	O75015	homo sapien
178	6	1.4	233	1	YCBG_BACSU	P42239	bacillus su
179	6	1.4	237	1	CD63_HUMAN	P08962	homo sapien
180	6	1.4	237	1	CD63_RABIT	Q28709	oryctolagus
181	6	1.4	237	1	YS76_CAEEL	Q09619	caenorhabdi
182	6	1.4	238	1	YEEN_ECOLI	P76351	escherichia
183	6	1.4	241	1	FADR_HAEIN	P44705	haemophilus
184	6	1.4	244	1	TRUA_MYCGE	P47428	mycoplasma
185	6	1.4	245	1	YLPA_YEREN	P27461	yersinia en
186	6	1.4	246	1	CFI_ARATH	P41088	arabidopsis
187	6	1.4	247	1	CAH_METTE	P40881	methanosarc
188	6	1.4	247	1	EFTS_SPIPL	P34828	spirulina p
189	6	1.4	247	1	PYG2_MASLA	P29732	mastigoclad
190	6	1.4	249	1	COBS_MYCTU	Q10397	mycobacteri
191	6	1.4	252	1	PMM_CANAL	P31353	candida alb
192	6	1.4	252	1	PSA6_YEAST	P21243	saccharomyc
193	6	1.4	252	1	UDP_SALTY	O33808	salmonella
194	6	1.4	253	1	UDP_KLEAE	O08444	klebsiella
195	6	1.4	254	1	FC3A_HUMAN	P08637	homo sapien
196	6	1.4	254	1	VSPB_SOYBN	P10743	glycine max
197	6	1.4	255	1	EI2B_DROME	P08761	drosophila
198	6	1.4	255	1	ETFB_HUMAN	P38117	homo sapien
199	6	1.4	255	1	HCD2_DROME	O18404	drosophila
200	6	1.4	255	1	HYCG_ECOLI	P16433	escherichia
201	6	1.4	255	1	TATC_AZOCH	P54085	azotobacter
202	6	1.4	256	1	YDH1_HSVSC	P22575	herpesvirus
203	6	1.4	257	1	ECH8_MYCLE	O07137	mycobacteri
204	6	1.4	257	1	ECH8_MYCTU	O53418	mycobacteri
205	6	1.4	257	1	ECHH_RHIME	Q52995	rhizobium m
206	6	1.4	257	1	OMPV_VIBCH	P06111	vibrio chol
207	6	1.4	257	1	PMM_SCHPO	Q9utj2	schizosacch
208	6	1.4	259	1	TRUA_METJA	Q59069	methanococc
209	6	1.4	262	1	S3AD_KLEPN	P08881	klebsiella
210	6	1.4	263	1	S3AD_ECOLI	P04826	escherichia
211	6	1.4	264	1	COLI_MACNE	P01201	m corticotr
212	6	1.4	264	1	TRPC_LACLA	Q01999	lactococcus
213	6	1.4	264	1	YXEM_BACSU	P54952	bacillus su
214	6	1.4	266	1	KLR8_MOUSE	Q60682	mus musculu
215	6	1.4	267	1	FTSQ_RICPR	Q9zds5	rickettsia
216	6	1.4	268	1	RAD_RAT	P55043	rattus norv
217	6	1.4	269	1	PRTM_BACSP	Q99405	bacillus sp
218	6	1.4	269	1	RAD_HUMAN	P55042	homo sapien
219	6	1.4	269	1	SUBB_BACLE	P29599	bacillus le
220	6	1.4	269	1	SUBS_BACLE	P29600	bacillus le
221	6	1.4	269	1	TRC1_STRCO	O68814	streptomyce
222	6	1.4	270	1	MER1_YEAST	P16523	saccharomyc
223	6	1.4	271	1	ECE1_CANAL	Q07730	candida alb
224	6	1.4	272	1	E2F6_MOUSE	O54917	mus musculu
225	6	1.4	273	1	ROC1_NICSY	Q08935	nicotiana s
226	6	1.4	274	1	Y975_TREPA	O83940	treponema p
227	6	1.4	276	1	KLKA_HUMAN	O43240	homo sapien
228	6	1.4	276	1	SSUC_BACSU	P40401	bacillus su
229	6	1.4	277	1	CYST_ECOLI	P16701	escherichia
230	6	1.4	279	1	DAPF_SYNY3	P74667	synechocyst
231	6	1.4	279	1	ROC1_NICPL	P49313	nicotiana p
232	6	1.4	279	1	UDP_KLEPN	P52671	klebsiella
233	6	1.4	279	1	Y144_MYCGE	P47390	mycoplasma
234	6	1.4	283	1	CXB1_HUMAN	P08034	homo sapien
235	6	1.4	283	1	CXB1_MOUSE	P28230	mus musculu
236	6	1.4	283	1	CXB1_RAT	P08033	rattus norv
237	6	1.4	284	1	CXB1_BOVIN	O18968	bos taurus
238	6	1.4	284	1	STX2_CAEEL	Q20574	caenorhabdi
239	6	1.4	285	1	PYR1_MASLA	P11398	mastigoclad
240	6	1.4	286	1	PYR1_ANASP	P07123	anabaena sp
241	6	1.4	286	1	PYR1_SYNEL	P50034	synechococc
242	6	1.4	290	1	ALF2_BACSU	P42420	bacillus su
243	6	1.4	290	1	ATPG_BUCAI	P57123	buchnera ap
244	6	1.4	290	1	VP38_HSVMG	P30023	marek's dis
245	6	1.4	290	1	YPUA_BACSU	P31847	bacillus su
246	6	1.4	290	1	YQI2_CAEEL	Q09504	caenorhabdi
247	6	1.4	291	1	C552_PSEST	P24037	pseudomonas
248	6	1.4	293	1	PARB_CAUCR	O05190	caulobacter
249	6	1.4	295	1	PEXA_PICAN	Q00940	pichia angu
250	6	1.4	295	1	YHYA_PSESN	Q01265	pseudomonas
251	6	1.4	297	1	BMRU_BACSU	P39074	bacillus su
252	6	1.4	297	1	STX4_HUMAN	Q12846	homo sapien

253	6	1.4	297	1	YNZ4_CAEEL	P45964	caenorhabdi
254	6	1.4	299	1	SC5D_HUMAN	O75845	homo sapien
255	6	1.4	301	1	FMDC_METBA	Q48943	methanosarc
256	6	1.4	302	1	ATPG_STRLI	P50007	streptomyce
257	6	1.4	302	1	PP12_DROME	P12982	drosophila
258	6	1.4	302	1	PP13_DROME	Q05547	drosophila
259	6	1.4	304	1	GAT1_CHICK	P17678	gallus gall
260	6	1.4	304	1	Y245_SYNY3	P72702	synechocyst
261	6	1.4	307	1	BRAD_PSEAE	P21627	pseudomonas
262	6	1.4	307	1	VP53_BPAPS	Q9t1p5	bacteriopha
263	6	1.4	308	1	Y880_METJA	Q58290	methanococc
264	6	1.4	309	1	MAT1_MOUSE	P51949	mus musculu
265	6	1.4	314	1	PIX1_HUMAN	P78337	homo sapien
266	6	1.4	315	1	O3A3_HUMAN	P47888	homo sapien
267	6	1.4	315	1	PIX1_MOUSE	P70314	mus musculu
268	6	1.4	317	1	MSHR_ALCAA	P56442	alces alces
269	6	1.4	317	1	MSHR_BOVIN	P47798	bos taurus
270	6	1.4	317	1	MSHR_CANFA	O77616	canis famil
271	6	1.4	317	1	MSHR_CAPCA	P56443	capreolus c
272	6	1.4	317	1	MSHR_CAPHI	P56444	capra hircu
273	6	1.4	317	1	MSHR_CEREL	P56445	cervus elap
274	6	1.4	317	1	MSHR_DAMDA	P56446	dama dama (
275	6	1.4	317	1	MSHR_OVIMO	P56447	ovibos mosc
276	6	1.4	317	1	MSHR_RANTA	P56448	rangifer ta
277	6	1.4	317	1	MSHR_SHEEP	O19037	ovis aries
278	6	1.4	317	1	MSHR_VULVU	Q29154	vulpes vulp
279	6	1.4	318	1	LPSA_BACNO	P39907	bacteroides
280	6	1.4	319	1	PTHB_ECOLI	P56580	escherichia
281	6	1.4	319	1	ROA1_MACMU	Q28521	macaca mula
282	6	1.4	319	1	ROA1_MOUSE	P49312	mus musculu
283	6	1.4	319	1	ROA1_RAT	P04256	rattus norv
284	6	1.4	321	1	ACCO_DIACA	P31528	dianthus ca
285	6	1.4	322	1	RM10_YEAST	P36520	saccharomyc
286	6	1.4	323	1	Y4LF_RHISN	P55546	rhizobium s
287	6	1.4	325	1	CYSK_CITLA	Q43317	citrullus l
288	6	1.4	325	1	CYSK_SPIOL	Q00834	spinacia ol
289	6	1.4	325	1	SBCD_BACSU	P23479	bacillus su
290	6	1.4	327	1	PP11_DROME	P48461	drosophila
291	6	1.4	328	1	FLIG_AQUAE	O66891	aquifex aeo
292	6	1.4	329	1	EXBB_PSEPU	Q05605	pseudomonas
293	6	1.4	330	1	FEPG_ECOLI	P23877	escherichia
294	6	1.4	330	1	YDEZ_ECOLI	P77651	escherichia
295	6	1.4	332	1	ILVC_STRCO	Q9z565	streptomyce
296	6	1.4	332	1	SRG3_CAEEL	P46572	caenorhabdi
297	6	1.4	334	1	RIR2_ASFB7	P42492	african swi
298	6	1.4	335	1	YJ89_PYRHO	O57713	pyrococcus
299	6	1.4	337	1	HAIR_DROME	P14003	drosophila
300	6	1.4	339	1	KMOS_RAT	P00539	rattus norv
301	6	1.4	340	1	AA2B_CHICK	O13076	gallus gall
302	6	1.4	340	1	UL20_HCMVA	P16758	human cytom
303	6	1.4	340	1	YDDR_ECOLI	P77308	escherichia
304	6	1.4	341	1	MOAA_BACSU	P39757	bacillus su
305	6	1.4	342	1	KMOS_MSVTS	P10421	myeloprolif
306	6	1.4	344	1	CHSA_PHANI	P48395	pharbitis n
307	6	1.4	344	1	HOLA_HAEIN	P43747	haemophilus
308	6	1.4	345	1	CEBB_HUMAN	P17676	homo sapien
309	6	1.4	346	1	YA0D_SCHPO	Q09681	schizosacch
310	6	1.4	347	1	VP11_WTV	P13094	wound tumor
311	6	1.4	349	1	25A1_PIG	Q29599	sus scrofa
312	6	1.4	349	1	OAS1_PIG	Q29599	sus scrofa
313	6	1.4	351	1	HRCA_MYCPN	P75351	mycoplasma
314	6	1.4	351	1	RFBB_XANCA	P55295	xanthomonas
315	6	1.4	351	1	YIIG_ECOLI	P32151	escherichia
316	6	1.4	352	1	CHSC_IPOPU	P48399	ipomoea pur
317	6	1.4	352	1	CLCE_PSEAE	O87612	pseudomonas
318	6	1.4	352	1	CLCE_PSESB	O30847	pseudomonas
319	6	1.4	352	1	YM63_CAEEL	P34524	caenorhabdi
320	6	1.4	353	1	TCPH_TETTH	P54410	tetrahymena
321	6	1.4	354	1	TFF1_ALCEU	P27137	alcaligenes
322	6	1.4	356	1	NTRB_RHOCA	P09431	rhodobacter
323	6	1.4	359	1	AC14_CHICK	P53033	gallus gall
324	6	1.4	359	1	DHOM_YEAST	P31116	saccharomyc
325	6	1.4	359	1	GCPE_HELPJ	Q9zl10	helicobacte
326	6	1.4	359	1	GCPE_HELPY	O25342	helicobacte
327	6	1.4	361	1	CHSA_IPOCO	P48393	ipomoea cor
328	6	1.4	361	1	DBPA_RAT	Q62764	rattus norv
329	6	1.4	361	1	PDA6_ARATH	O22263	arabidopsis
330	6	1.4	361	1	YJGQ_ECOLI	P39341	escherichia
331	6	1.4	362	1	CHSA_IPOPL	P48400	ipomoea pla
332	6	1.4	362	1	CHSA_IPOTF	P48401	ipomoea tri
333	6	1.4	362	1	KLF1_HUMAN	Q13351	homo sapien
334	6	1.4	363	1	AROC_PSEAE	Q9i344	pseudomonas
335	6	1.4	363	1	YGCJ_ECOLI	Q46899	escherichia
336	6	1.4	365	1	CMLE_NEUCR	P38677	neurospora
337	6	1.4	367	1	CCR3_MOUSE	O88410	mus musculu
338	6	1.4	367	1	Y797_METJA	Q58207	methanococc
339	6	1.4	368	1	CCR3_HUMAN	P49682	homo sapien
340	6	1.4	368	1	ZN24_HUMAN	P17028	homo sapien
341	6	1.4	371	1	ROA1_HUMAN	P09651	homo sapien
342	6	1.4	373	1	LEU3_CANAL	P87186	candida alb
343	6	1.4	374	1	FLA1_BARBA	P35633	bartonella
344	6	1.4	374	1	KMOS_MSVMH	P07331	moloney mur
345	6	1.4	375	1	NY4R_MOUSE	Q61041	mus musculu
346	6	1.4	376	1	GBAS_LYMST	P30684	lymnaea sta
347	6	1.4	376	1	KMOS_MSVMM	P00537	moloney mur
348	6	1.4	376	1	PTI6_HUMAN	P35237	homo sapien
349	6	1.4	377	1	HSF7_ARATH	Q9t0d3	arabidopsis
350	6	1.4	377	1	RECA_STRVL	P48295	streptomyce
351	6	1.4	378	1	ELYA_BACSP	P20724	bacillus sp
352	6	1.4	378	1	HAIR_DROVI	P29303	drosophila
353	6	1.4	380	1	ELYA_BACAO	P27693	bacillus al
354	6	1.4	380	1	ELYA_BACCS	P41362	bacillus cl
355	6	1.4	381	1	NIFV_KLEPN	P05345	klebsiella
356	6	1.4	381	1	YNU6_CAEEL	P50444	caenorhabdi
357	6	1.4	382	1	HUGA_APIME	Q08169	apis mellif
358	6	1.4	382	1	PARG_HUMAN	P51888	homo sapien
359	6	1.4	384	1	GBA1_LUPLU	Q40224	lupinus lut
360	6	1.4	384	1	SDC3_HUMAN	O75056	homo sapien
361	6	1.4	386	1	YMK7_YEAST	Q03760	saccharomyc
362	6	1.4	387	1	GAT6_CHICK	P43693	gallus gall
363	6	1.4	387	1	INVO_TARBA	P24711	tarsius ban
364	6	1.4	392	1	YAY1_SCHPO	Q10209	schizosacch
365	6	1.4	393	1	AAT_BACST	Q59228	bacillus st
366	6	1.4	393	1	SRPC_SYNP7	Q55027	synechococc
367	6	1.4	394	1	Z193_HUMAN	O15535	homo sapien
368	6	1.4	396	1	CATE_HUMAN	P14091	homo sapien
369	6	1.4	396	1	CHSA_IPOPU	P48397	ipomoea pur
370	6	1.4	396	1	DCOP_SORMA	P78748	sordaria ma
371	6	1.4	397	1	DCOP_NEUCR	P05035	neurospora
372	6	1.4	397	1	PGK_XANFL	P50314	xanthobacte
373	6	1.4	397	1	TYRB_ECOLI	P04693	escherichia
374	6	1.4	397	1	YKNZ_BACSU	O31712	bacillus su
375	6	1.4	399	1	CXA8_CHICK	P36381	gallus gall
376	6	1.4	400	1	CRT1_BOVIN	P52193	bos taurus
377	6	1.4	401	1	AMPC_PSYIM	O05465	psychrobact
378	6	1.4	403	1	ARRS_MOUSE	P20443	mus musculu
379	6	1.4	403	1	ARRS_RAT	P15887	rattus norv
380	6	1.4	403	1	YA24_METJA	Q58430	methanococc
381	6	1.4	404	1	LMP1_EBVC	P29362	epstein-bar
382	6	1.4	405	1	CARP_YEAST	P07267	saccharomyc
383	6	1.4	405	1	YLF2_YEAST	P38746	saccharomyc
384	6	1.4	405	1	YS63_CAEEL	Q09379	caenorhabdi
385	6	1.4	406	1	TRPB_RHIET	P56929	rhizobium e
386	6	1.4	407	1	NUDM_DROME	P91929	drosophila
387	6	1.4	407	1	SYY_THIFE	P41256	thiobacillu
388	6	1.4	411	1	RAPS_TORCA	P09108	torpedo cal
389	6	1.4	411	1	SYY_RICPR	Q9zcz4	rickettsia
390	6	1.4	411	1	VGLM_HSVBC	P52370	bovine herp
391	6	1.4	411	1	Y552_STRCO	P40180	streptomyce
392	6	1.4	412	1	CAR2_CANPA	P32950	candida par
393	6	1.4	412	1	YNP1_CAEEL	P34554	caenorhabdi
394	6	1.4	415	1	EUTH_SALTY	P41796	salmonella
395	6	1.4	415	1	MNTH_MYCBO	O69443	mycobacteri
396	6	1.4	415	1	ZP3_RABIT	P48833	oryctolagus
397	6	1.4	416	1	CRTC_MOUSE	P14211	mus musculu
398	6	1.4	416	1	CRTC_RAT	P18418	rattus norv

399	6	1.4	417	1	CRTC_HUMAN	P27797	homo sapien
400	6	1.4	417	1	DHMH_PARDE	P29894	paracoccus
401	6	1.4	417	1	Z232_HUMAN	Q9uny5	homo sapien
402	6	1.4	418	1	CRTC_RABIT	P15253	oryctolagus
403	6	1.4	419	1	VEGC_HUMAN	P49767	homo sapien
404	6	1.4	420	1	EF1A_HALMA	P16018	haloarcula
405	6	1.4	421	1	EF1A_HALHA	P48863	halobacteri
406	6	1.4	423	1	MBI2_YEAST	P03873	saccharomyc
407	6	1.4	423	1	POLS_SINDW	P11259	sindbis vir
408	6	1.4	424	1	SYY_MYCTU	O33191	mycobacteri
409	6	1.4	425	1	HE47_CAEEL	Q18212	caenorhabdi
410	6	1.4	427	1	FLT1_HUMAN	O75955	homo sapien
411	6	1.4	428	1	EF1A_PYRHO	O59153	pyrococcus
412	6	1.4	428	1	FLT1_MOUSE	O08917	mus musculu
413	6	1.4	428	1	FLT1_RAT	Q9z1e1	rattus norv
414	6	1.4	428	1	MNTH_MYCTU	O05916	mycobacteri
415	6	1.4	428	1	PROA_TREPA	P74935	treponema p
416	6	1.4	429	1	YFAV_ECOLI	P76470	escherichia
417	6	1.4	430	1	EF1A_PYRWO	P26751	pyrococcus
418	6	1.4	431	1	NFIX_CHICK	Q90932	gallus gall
419	6	1.4	432	1	APRE_PSEAE	Q03025	pseudomonas
420	6	1.4	432	1	YF10_MYCTU	P71789	mycobacteri
421	6	1.4	433	1	UROK_BOVIN	Q05589	bos taurus
422	6	1.4	435	1	R114_YEAST	P32841	saccharomyc
423	6	1.4	435	1	YYBO_BACSU	P37489	bacillus su
424	6	1.4	435	1	YYS3_CAEEL	P49191	caenorhabdi
425	6	1.4	436	1	EPSE_BURSO	Q45411	burkholderi
426	6	1.4	437	1	PEPG_LACDL	P94869	lactobacill
427	6	1.4	439	1	Y441_BUCAI	P57516	buchnera ap
428	6	1.4	440	1	LCAT_RABIT	P53761	oryctolagus
429	6	1.4	441	1	GSA_PROFR	Q06774	propionibac
430	6	1.4	442	1	HPPD_DAUCA	O23920	daucus caro
431	6	1.4	446	1	DCUB_ECOLI	P14409	escherichia
432	6	1.4	447	1	EF1A_PEA	Q41011	pisum sativ
433	6	1.4	447	1	EF1A_SOYBN	P25698	glycine max
434	6	1.4	447	1	EF1A_VICFA	O24534	vicia faba
435	6	1.4	449	1	EF11_DAUCA	P29521	daucus caro
436	6	1.4	449	1	MB21_ARATH	O80998	arabidopsis
437	6	1.4	450	1	CBPA_SYNP7	P39660	synechococc
438	6	1.4	452	1	PTCC_ECOLI	P17334	escherichia
439	6	1.4	453	1	YWZ5_CAEEL	Q11095	caenorhabdi
440	6	1.4	454	1	YUA6_CAEEL	P52715	caenorhabdi
441	6	1.4	457	1	DBDR_XENLA	P42290	xenopus lae
442	6	1.4	458	1	NIFN_AZOVI	P10336	azotobacter
443	6	1.4	461	1	ATOC_ECOLI	Q06065	escherichia
444	6	1.4	461	1	CDS1_HUMAN	Q92903	h phosphati
445	6	1.4	461	1	EF1A_ARTSA	P02993	artemia sal
446	6	1.4	462	1	CATC_MOUSE	P97821	mus musculu
447	6	1.4	462	1	EF1A_BRARE	Q92005	brachydanio
448	6	1.4	464	1	CDK8_HUMAN	P49336	homo sapien
449	6	1.4	466	1	IMP2_DROME	P29681	drosophila
450	6	1.4	466	1	SX10_HUMAN	P56693	homo sapien
451	6	1.4	466	1	SX10_MOUSE	Q04888	mus musculu
452	6	1.4	466	1	SX10_RAT	O55170	rattus norv
453	6	1.4	466	1	SYN_BUCAI	P57441	buchnera ap
454	6	1.4	467	1	B1AR_BOVIN	Q9tt96	bos taurus
455	6	1.4	467	1	E2BD_SCHPO	Q09924	schizosacch
456	6	1.4	468	1	B1AR_PIG	Q28998	sus scrofa
457	6	1.4	468	1	PPAD_YEAST	P52290	saccharomyc
458	6	1.4	468	1	SP5R_BACSU	P37875	bacillus su
459	6	1.4	473	1	B1AR_CANFA	P79148	canis famil
460	6	1.4	474	1	B1AR_FELCA	Q9tst6	felis silve
461	6	1.4	475	1	DLDH_VIBPA	O50286	vibrio para
462	6	1.4	475	1	RP54_THIFE	P24695	thiobacillu
463	6	1.4	477	1	B1AR_HUMAN	P08588	homo sapien
464	6	1.4	477	1	DLDH_TRYCR	P90597	trypanosoma
465	6	1.4	478	1	EA6_ARATH	Q06915	arabidopsis
466	6	1.4	480	1	B1AR_MACMU	P47899	macaca mula
467	6	1.4	481	1	LMRA_STRLN	P46104	streptomyce
468	6	1.4	481	1	Y209_ARCFU	O30030	archaeoglob
469	6	1.4	481	1	YEAV_ECOLI	P76252	escherichia
470	6	1.4	482	1	PUR8_SCHPO	O60105	schizosacch
471	6	1.4	483	1	FIBA_BOVIN	P02672	bos taurus
472	6	1.4	483	1	MURE_CHLTR	O84271	chlamydia t
473	6	1.4	483	1	XYLB_STAXY	P27155	staphylococ
474	6	1.4	485	1	OPRM_PSEAE	Q51487	pseudomonas
475	6	1.4	485	1	Z165_HUMAN	P49910	homo sapien
476	6	1.4	486	1	EXU_DROVI	Q24747	drosophila
477	6	1.4	487	1	Y442_MYCTU	P42611	mycobacteri
478	6	1.4	488	1	ASTD_PSEAE	O50174	pseudomonas
479	6	1.4	489	1	C128_MYCTU	Q59572	mycobacteri
480	6	1.4	490	1	TMS2_MOUSE	Q9jiq8	mus musculu
481	6	1.4	496	1	C4AE_DROME	O46054	drosophila
482	6	1.4	496	1	CAMA_HUMAN	P21941	homo sapien
483	6	1.4	496	1	KPR5_YEAST	Q12265	saccharomyc
484	6	1.4	498	1	GATA_EMENI	P14010	emericella
485	6	1.4	498	1	SYC_DEIRA	Q9rtt6	deinococcus
486	6	1.4	499	1	GLPK_SULSO	P95907	sulfolobus
487	6	1.4	499	1	RRG_BRARE	Q91392	brachydanio
488	6	1.4	500	1	YJEM_ECOLI	P39282	escherichia
489	6	1.4	501	1	LYSI_CORGL	P35865	corynebacte
490	6	1.4	501	1	SYK_HELPJ	Q9zmp8	helicobacte
491	6	1.4	501	1	SYK_HELPY	P56126	helicobacte
492	6	1.4	501	1	TX18_HUMAN	O95935	homo sapien
493	6	1.4	504	1	SPAS_MOUSE	Q9qyy8	mus musculu
494	6	1.4	504	1	YHV9_YEAST	P38854	saccharomyc
495	6	1.4	505	1	C8O1_HUMAN	Q9y236	homo sapien
496	6	1.4	505	1	PRI2_MOUSE	P33610	mus musculu
497	6	1.4	507	1	C7C3_ARATH	O64638	arabidopsis
498	6	1.4	507	1	RECN_CAMJE	Q9phm5	campylobact
499	6	1.4	508	1	ASPR_HORVU	P42210	hordeum vul
500	6	1.4	509	1	G6PD_ANASP	P48992	anabaena sp
501	6	1.4	509	1	G6PD_SYNY3	P73411	synechocyst
502	6	1.4	509	1	NCAP_SV5	Q88435	simian viru
503	6	1.4	509	1	PRI2_HUMAN	P49643	homo sapien
504	6	1.4	510	1	ACHG_XENLA	P05376	xenopus lae
505	6	1.4	510	1	NOA1_HUMAN	P51513	homo sapien
506	6	1.4	511	1	Y4JD_RHISN	P55504	rhizobium s
507	6	1.4	513	1	SYS_METTH	O27194	methanobact
508	6	1.4	515	1	CP12_RABIT	P00187	oryctolagus
509	6	1.4	518	1	KKL6_YEAST	P28708	saccharomyc
510	6	1.4	520	1	DPY1_CAEEL	Q18677	caenorhabdi
511	6	1.4	521	1	YAV8_SCHPO	Q10177	schizosacch
512	6	1.4	524	1	G6PD_SYNP7	P29686	synechococc
513	6	1.4	524	1	P60_LISWE	Q01839	listeria we
514	6	1.4	525	1	CIK3_RAT	P15384	rattus norv
515	6	1.4	526	1	MUTL_CAUCR	Q9rp66	caulobacter
516	6	1.4	528	1	CIK3_MOUSE	P16390	mus musculu
517	6	1.4	529	1	MBA1_ECOLI	P08098	escherichia
518	6	1.4	530	1	HLY_HAL17	P29143	halophilic
519	6	1.4	532	1	VSI3_TRYBB	P26328	trypanosoma
520	6	1.4	533	1	GTR2_CHICK	Q90592	gallus gall
521	6	1.4	533	1	NCR1_RAT	Q9wub5	rattus norv
522	6	1.4	534	1	CN1B_BOVIN	Q01061	bos taurus
523	6	1.4	536	1	CN1B_HUMAN	Q01064	homo sapien
524	6	1.4	536	1	SYEM_YEAST	P48525	saccharomyc
525	6	1.4	542	1	IMA1_SCHPO	O14063	schizosacch
526	6	1.4	544	1	ODP2_ACHLA	P35489	acholeplasm
527	6	1.4	545	1	HTR5_HALN1	Q48318	halobacteri
528	6	1.4	545	1	SYN_CAEEL	Q19722	caenorhabdi
529	6	1.4	547	1	ODP2_PSEAE	Q59638	pseudomonas
530	6	1.4	548	1	SYN_BRUMA	P10723	brugia mala
531	6	1.4	550	1	PUTX_EMENI	P18696	emericella
532	6	1.4	551	1	SYE_ARCFU	O29979	archaeoglob
533	6	1.4	552	1	MTJ1_MOUSE	Q61712	mus musculu
534	6	1.4	553	1	AAAX_MOUSE	P51912	mus musculu
535	6	1.4	555	1	ZF38_MOUSE	Q07231	mus musculu
536	6	1.4	556	1	PLD_STRAT	Q53728	streptomyce
537	6	1.4	558	1	TCPH_TETPY	P54409	tetrahymena
538	6	1.4	559	1	FLIF_SALTY	P15928	salmonella
539	6	1.4	560	1	YML2_ARATH	O22757	arabidopsis
540	6	1.4	562	1	GUN1_ACICE	P54583	acidothermu
541	6	1.4	563	1	DP3X_BACSU	P09122	bacillus su
542	6	1.4	563	1	PK66_PLAKU	P21303	plasmodium
543	6	1.4	565	1	ES10_RAT	P16303	rattus norv
544	6	1.4	566	1	MXID_SHISO	Q55293	shigella so

545	6	1.4	567	1	KDGE_HUMAN	P52429	homo sapien
546	6	1.4	572	1	NU5M_LOCMI	Q36428	locusta mig
547	6	1.4	575	1	I10R_MOUSE	Q61727	mus musculu
548	6	1.4	576	1	PRIM_LEGPN	P71481	legionella
549	6	1.4	576	1	YIT0_YEAST	P40568	saccharomyc
550	6	1.4	577	1	ALG8_YEAST	P40351	saccharomyc
551	6	1.4	577	1	STS_RAT	P15589	rattus norv
552	6	1.4	579	1	YHZ4_YEAST	P38880	saccharomyc
553	6	1.4	580	1	IPAB_SHIDY	Q03945	shigella dy
554	6	1.4	580	1	IPAB_SHIFL	P18011	shigella fl
555	6	1.4	583	1	SYR_AQUAE	O67068	aquifex aeo
556	6	1.4	583	1	YF21_MYCTU	Q50586	mycobacteri
557	6	1.4	584	1	PME1_CITSI	O04886	citrus sine
558	6	1.4	585	1	GLPD_MYCLE	P53435	mycobacteri
559	6	1.4	585	1	YAOH_SCHPO	Q10096	schizosacch
560	6	1.4	590	1	THIC_BACSU	P45740	bacillus su
561	6	1.4	592	1	PT1_ALCEU	P23536	alcaligenes
562	6	1.4	593	1	Y917_MYCTU	O05909	mycobacteri
563	6	1.4	594	1	DHBK_LYCES	O04059	lycopersico
564	6	1.4	594	1	RBJK_DROME	P28159	drosophila
565	6	1.4	602	1	LEPA_CHLPN	Q9z814	chlamydia p
566	6	1.4	602	1	LEPA_CHLTR	O84067	chlamydia t
567	6	1.4	603	1	LCB2_SCHPO	Q09925	schizosacch
568	6	1.4	606	1	GAB_DROME	P25123	drosophila
569	6	1.4	609	1	YL01_YEAST	Q06053	saccharomyc
570	6	1.4	610	1	THIC_RHIET	O34291	rhizobium e
571	6	1.4	612	1	Y525_PSEAE	Q51484	pseudomonas
572	6	1.4	612	1	YMY8_YEAST	Q03153	saccharomyc
573	6	1.4	613	1	ML1X_HUMAN	Q13585	homo sapien
574	6	1.4	619	1	IF2_MYCGE	P47388	mycoplasma
575	6	1.4	623	1	OGP_PAPAN	P36718	papio anubi
576	6	1.4	623	1	PNT1_DROME	P51022	drosophila
577	6	1.4	625	1	TBP2_HAEIN	P44971	haemophilus
578	6	1.4	629	1	ODP2_ECOLI	P06959	escherichia
579	6	1.4	638	1	OAR1_LYMST	O77408	lymnaea sta
580	6	1.4	639	1	TETM_STRLI	Q02652	streptomyce
581	6	1.4	642	1	YB65_SCHPO	Q09746	schizosacch
582	6	1.4	645	1	IME2_YEAST	P32581	saccharomyc
583	6	1.4	645	1	SUBE_BACSU	P16396	bacillus su
584	6	1.4	646	1	PLAP_MOUSE	P27612	mus musculu
585	6	1.4	646	1	VP40_HSVEB	P28936	equine herp
586	6	1.4	647	1	PLAP_RAT	P54319	rattus norv
587	6	1.4	651	1	YHJK_ECOLI	P37649	escherichia
588	6	1.4	654	1	TFE2_HUMAN	P15923	homo sapien
589	6	1.4	656	1	YK79_MYCTU	Q10687	mycobacteri
590	6	1.4	660	1	VG06_BPT4	P19060	bacteriopha
591	6	1.4	661	1	ACSA_COPCI	O13440	coprinus ci
592	6	1.4	665	1	Y176_METTH	O26278	methanobact
593	6	1.4	667	1	SIM2_HUMAN	Q14190	homo sapien
594	6	1.4	670	1	INV1_MAIZE	P49175	zea mays (m
595	6	1.4	677	1	FLHA_BACSU	P35620	bacillus su
596	6	1.4	678	1	OGP_HUMAN	Q12889	homo sapien
597	6	1.4	678	1	YF48_MYCTU	Q10778	mycobacteri
598	6	1.4	680	1	GAG_SCVLA	P32503	saccharomyc
599	6	1.4	687	1	TRFE_ONCKI	P79815	oncorhynchu
600	6	1.4	690	1	IF2_THEMA	Q9wzn3	thermotoga
601	6	1.4	690	1	TOP1_BACHD	Q9ka23	bacillus ha
602	6	1.4	692	1	DNK1_SYNY3	Q55154	synechocyst
603	6	1.4	692	1	RFX2_MOUSE	P48379	mus musculu
604	6	1.4	695	1	YD38_YEAST	Q05497	saccharomyc
605	6	1.4	696	1	YO12_CAEEL	P34668	caenorhabdi
606	6	1.4	698	1	YB06_YEAST	P38283	saccharomyc
607	6	1.4	700	1	FLHA_CAUCR	Q03845	caulobacter
608	6	1.4	701	1	Y363_BUCAI	P57444	buchnera ap
609	6	1.4	703	1	UL17_HSV11	P10201	herpes simp
610	6	1.4	706	1	GYS_NEUCR	O93869	neurospora
611	6	1.4	712	1	NRDD_ECOLI	P28903	escherichia
612	6	1.4	718	1	PNT2_DROME	P51023	drosophila
613	6	1.4	720	1	AGAL_STRMU	P27756	streptococc
614	6	1.4	721	1	Y781_METJA	Q58191	methanococc
615	6	1.4	722	1	P85B_MOUSE	O08908	mus musculu
616	6	1.4	722	1	P85B_RAT	Q63788	rattus norv
617	6	1.4	723	1	RFX2_HUMAN	P48378	homo sapien
618	6	1.4	723	1	SSRP_DROME	Q05344	drosophila
619	6	1.4	724	1	P85B_BOVIN	P23726	bos taurus
620	6	1.4	727	1	IF2M_HUMAN	P46199	homo sapien
621	6	1.4	727	1	VP4_RDVA	P22474	rice dwarf
622	6	1.4	727	1	VP4_RDVF	Q85436	rice dwarf
623	6	1.4	728	1	P85B_HUMAN	O00459	homo sapien
624	6	1.4	733	1	AGA1_PEDPE	P43467	pediococcus
625	6	1.4	735	1	ADDA_MOUSE	Q9qyc0	mus musculu
626	6	1.4	735	1	ADDA_RAT	Q63028	rattus norv
627	6	1.4	738	1	PLAP_HUMAN	Q9y263	homo sapien
628	6	1.4	740	1	FAS_PNECA	P29251	p folic aci
629	6	1.4	741	1	ERF2_PICPI	P23637	pichia pinu
630	6	1.4	742	1	ZFA_MOUSE	P23607	mus musculu
631	6	1.4	747	1	CLC4_MOUSE	Q61418	mus musculu
632	6	1.4	747	1	CLC4_RAT	P51794	rattus norv
633	6	1.4	748	1	HEPA_HSVBC	P52374	bovine herp
634	6	1.4	749	1	SWAP_CAEEL	Q10580	caenorhabdi
635	6	1.4	756	1	PID1_HUMAN	P51178	homo sapien
636	6	1.4	758	1	CLPA_ECOLI	P15716	escherichia
637	6	1.4	760	1	CLC4_HUMAN	P51793	homo sapien
638	6	1.4	767	1	PUB1_SCHPO	Q92462	schizosacch
639	6	1.4	767	1	YNW4_YEAST	P53866	saccharomyc
640	6	1.4	770	1	PCK7_MOUSE	Q61139	mus musculu
641	6	1.4	771	1	CTPG_MYCTU	Q10866	mycobacteri
642	6	1.4	773	1	MSL2_DROME	P50534	drosophila
643	6	1.4	775	1	MCM5_YEAST	P29496	saccharomyc
644	6	1.4	780	1	AFG2_YEAST	P32794	saccharomyc
645	6	1.4	783	1	PCK7_RAT	Q62849	rattus norv
646	6	1.4	785	1	PCK7_HUMAN	Q16549	homo sapien
647	6	1.4	785	1	VD05_VACCC	P21010	vaccinia vi
648	6	1.4	785	1	VD05_VACCV	P04305	vaccinia vi
649	6	1.4	785	1	VD05_VARV	P33069	variola vir
650	6	1.4	788	1	PUR2_YARLI	Q99148	y bifunctio
651	6	1.4	790	1	QUIA_XANCJ	Q9xd78	xanthomonas
652	6	1.4	794	1	TRK1_LYMST	O76997	lymnaea sta
653	6	1.4	794	1	Z151_MOUSE	Q60821	mus musculu
654	6	1.4	799	1	ZFX1_MOUSE	P17011	mus musculu
655	6	1.4	800	1	ZFX_BOVIN	O62836	bos taurus
656	6	1.4	801	1	FGR3_MOUSE	Q61851	mus musculu
657	6	1.4	801	1	ZFY_HUMAN	P08048	homo sapien
658	6	1.4	803	1	Z151_HUMAN	Q13105	homo sapien
659	6	1.4	805	1	IF2_AQUAE	O67825	aquifex aeo
660	6	1.4	805	1	SUS1_TULGE	Q41608	tulipa gesn
661	6	1.4	805	1	ZFX_HUMAN	P17010	homo sapien
662	6	1.4	806	1	FGR3_HUMAN	P22607	homo sapien
663	6	1.4	806	1	LON_RHIME	O69177	rhizobium m
664	6	1.4	812	1	LON_BRUAB	O52605	brucella ab
665	6	1.4	815	1	NAH1_HUMAN	P19634	homo sapien
666	6	1.4	817	1	PPSA_PYRFU	P42850	pyrococcus
667	6	1.4	820	1	TRBE_AGRTU	P54910	agrobacteri
668	6	1.4	829	1	GYRA_CLOAB	P94605	clostridium
669	6	1.4	829	1	IF2_HAEIN	P44323	haemophilus
670	6	1.4	830	1	PKN2_MYXXA	P54736	myxococcus
671	6	1.4	830	1	PRLR_COLLI	Q90374	columba liv
672	6	1.4	830	1	VPP3_HUMAN	Q13488	h vacuolar
673	6	1.4	834	1	MSH5_HUMAN	O43196	homo sapien
674	6	1.4	838	1	GYRA_MYCTU	Q07702	mycobacteri
675	6	1.4	839	1	ZFX2_MOUSE	P17012	mus musculu
676	6	1.4	842	1	EF2_YEAST	P32324	saccharomyc
677	6	1.4	842	1	GYRA_MYCSM	P48354	mycobacteri
678	6	1.4	842	1	IF2_TREPA	O83861	treponema p
679	6	1.4	848	1	TOP1_BORBU	O51768	borrelia bu
680	6	1.4	852	1	RA54_SCHPO	P41410	schizosacch
681	6	1.4	854	1	SYV_THETH	P96142	thermus aqu
682	6	1.4	855	1	CNRC_BOVIN	P16586	bos taurus
683	6	1.4	860	1	LDLR_HUMAN	P01130	homo sapien
684	6	1.4	863	1	YP67_CAEEL	Q09216	caenorhabdi
685	6	1.4	867	1	EF2_BLAHO	Q17152	blastocysti
686	6	1.4	868	1	TOP1_HAEIN	P43012	haemophilus
687	6	1.4	868	1	VAV2_MOUSE	Q60992	mus musculu
688	6	1.4	870	1	Y563_HUMAN	O60309	homo sapien
689	6	1.4	872	1	SCD1_SCHPO	P40995	schizosacch
690	6	1.4	873	1	FIMC_BORPE	P33410	bordetella

691	6	1.4	874	1	SLAP_BACLI	P49052	bacillus li
692	6	1.4	876	1	TOP1_VIBCH	Q9krb2	vibrio chol
693	6	1.4	877	1	DPO1_LACLC	O32801	lactococcus
694	6	1.4	878	1	VAV2_HUMAN	P52735	homo sapien
695	6	1.4	880	1	ENV_SIVML	P11267	simian immu
696	6	1.4	881	1	ENV_SIVMK	P05884	simian immu
697	6	1.4	889	1	ENV_SIVSP	P19503	simian immu
698	6	1.4	890	1	IF2_ECOLI	P02995	escherichia
699	6	1.4	890	1	YCF2_MESVI	Q9mup8	mesostigma
700	6	1.4	892	1	AAC1_HUMAN	P12814	homo sapien
701	6	1.4	892	1	AAC1_RAT	Q9z1p2	rattus norv
702	6	1.4	892	1	ATX7_HUMAN	O15265	homo sapien
703	6	1.4	892	1	IF2_SALTY	Q9zf31	salmonella
704	6	1.4	893	1	AAC1_CHICK	P05094	gallus gall
705	6	1.4	894	1	WPRA_BACSU	P54423	bacillus su
706	6	1.4	896	1	BOSS_DROME	P22815	drosophila
707	6	1.4	896	1	IF2_KLEOX	Q9zf28	klebsiella
708	6	1.4	897	1	IF2_ENTCL	Q9zf25	enterobacte
709	6	1.4	898	1	FAS2_SCHAM	P22648	schistocerc
710	6	1.4	898	1	NIA_BETVE	P27783	betula verr
711	6	1.4	900	1	IF2_MYCTU	P71613	mycobacteri
712	6	1.4	903	1	VGLB_HSV1F	P06436	herpes simp
713	6	1.4	904	1	VGLB_HSV11	P10211	herpes simp
714	6	1.4	904	1	VGLB_HSV1P	P08665	herpes simp
715	6	1.4	915	1	SMOO_HUMAN	P53814	homo sapien
716	6	1.4	916	1	RTJK_DROME	P21328	drosophila
717	6	1.4	917	1	IF2_PROVU	Q9zf22	proteus vul
718	6	1.4	920	1	MML7_MYCTU	P96289	mycobacteri
719	6	1.4	921	1	NAC2_HUMAN	Q9upr5	homo sapien
720	6	1.4	921	1	NAC2_RAT	P48768	rattus norv
721	6	1.4	929	1	KDGZ_RAT	O08560	rattus norv
722	6	1.4	937	1	ODO1_BACSU	P23129	bacillus su
723	6	1.4	943	1	NFC1_HUMAN	O95644	homo sapien
724	6	1.4	946	1	YIA2_YEAST	P40559	saccharomyc
725	6	1.4	952	1	YH19_RHOCA	Q02998	rhodobacter
726	6	1.4	954	1	M3KA_HUMAN	Q02779	homo sapien
727	6	1.4	959	1	SYAC_SCHPO	O13914	schizosacch
728	6	1.4	960	1	CAP1_SORBI	P29195	sorghum bic
729	6	1.4	960	1	CAP2_MESCR	P16097	mesembryant
730	6	1.4	961	1	TSP4_HUMAN	P35443	homo sapien
731	6	1.4	962	1	YA4E_SCHPO	Q09731	schizosacch
732	6	1.4	966	1	MMLB_MYCTU	O53653	mycobacteri
733	6	1.4	966	1	ST10_MOUSE	O55098	mus musculu
734	6	1.4	967	1	CAP1_SOYBN	Q02909	glycine max
735	6	1.4	968	1	ST10_HUMAN	O94804	homo sapien
736	6	1.4	977	1	EPA2_MOUSE	Q03145	mus musculu
737	6	1.4	980	1	YRS5_CAEEL	Q09349	caenorhabdi
738	6	1.4	982	1	YS96_CAEEL	Q09965	caenorhabdi
739	6	1.4	986	1	GUNZ_CLOSR	P23659	clostridium
740	6	1.4	994	1	SYV_XYLFA	Q9ph12	xylella fas
741	6	1.4	1006	1	EPB6_HUMAN	O15197	homo sapien
742	6	1.4	1030	1	Y075_MYCPN	P75556	mycoplasma
743	6	1.4	1036	1	P200_MYCPN	P75211	mycoplasma
744	6	1.4	1039	1	MSL1_DROME	P50535	drosophila
745	6	1.4	1048	1	AGO1_ARATH	O04379	arabidopsis
746	6	1.4	1057	1	SPS1_CITUN	O22060	citrus unsh
747	6	1.4	1068	1	SPS_MAIZE	P31927	zea mays (m
748	6	1.4	1079	1	MY1A_MOUSE	P46735	mus musculu
749	6	1.4	1085	1	CUT7_SCHPO	P24339	schizosacch
750	6	1.4	1101	1	FURL_DROME	P30430	drosophila
751	6	1.4	1121	1	WDR6_HUMAN	Q9nnw5	homo sapien
752	6	1.4	1141	1	CN3A_HUMAN	Q14432	homo sapien
753	6	1.4	1147	1	NRDC_HUMAN	O43847	homo sapien
754	6	1.4	1151	1	ITA1_HUMAN	P56199	homo sapien
755	6	1.4	1173	1	ATC2_YEAST	P38929	saccharomyc
756	6	1.4	1173	1	ATCR_HUMAN	Q16720	homo sapien
757	6	1.4	1176	1	ATCP_RAT	P11505	rattus norv
758	6	1.4	1180	1	ITA1_RAT	P18614	rattus norv
759	6	1.4	1182	1	RTP2_TRYBG	P15594	trypanosoma
760	6	1.4	1185	1	MAPX_DROME	P23226	drosophila
761	6	1.4	1193	1	RPOB_BACSU	P37870	bacillus su
762	6	1.4	1198	1	ATCQ_HUMAN	Q01814	homo sapien
763	6	1.4	1198	1	ATCQ_RAT	P11506	rattus norv
764	6	1.4	1202	1	NOS3_HUMAN	P29474	homo sapien
765	6	1.4	1204	1	NOS3_PIG	Q28969	sus scrofa
766	6	1.4	1205	1	YAV9_SCHPO	Q10178	schizosacch
767	6	1.4	1209	1	THR_DROME	P42286	drosophila
768	6	1.4	1215	1	BGCN_DROME	Q9w1i2	drosophila
769	6	1.4	1220	1	ATCP_HUMAN	P20020	homo sapien
770	6	1.4	1220	1	ATCP_PIG	P23220	sus scrofa
771	6	1.4	1220	1	ATCP_RABIT	Q00804	oryctolagus
772	6	1.4	1224	1	RPB2_YEAST	P08518	saccharomyc
773	6	1.4	1228	1	ALA5_ARATH	Q9sgg3	arabidopsis
774	6	1.4	1229	1	KPB2_FUGRU	Q9w6r1	fugu rubrip
775	6	1.4	1238	1	DPOG_MOUSE	P54099	mus musculu
776	6	1.4	1245	1	POLS_SINDV	P03316	sindbis vir
777	6	1.4	1246	1	YMV2_CAEEL	P34504	caenorhabdi
778	6	1.4	1248	1	CYA8_RAT	P40146	rattus norv
779	6	1.4	1249	1	CYA8_MOUSE	P97490	mus musculu
780	6	1.4	1251	1	CYA8_HUMAN	P40145	homo sapien
781	6	1.4	1264	1	YGN3_YEAST	P53125	saccharomyc
782	6	1.4	1265	1	DYNA_DROME	P13496	drosophila
783	6	1.4	1270	1	DYNA_HUMAN	Q14203	homo sapien
784	6	1.4	1273	1	GYRA_MYCLE	Q57532	mycobacteri
785	6	1.4	1310	1	ICP4_VZVD	P09310	varicella-z
786	6	1.4	1323	1	NME4_MOUSE	Q03391	mus musculu
787	6	1.4	1336	1	MAM1_SCHPO	P78966	schizosacch
788	6	1.4	1342	1	XDH_DROPS	P22811	drosophila
789	6	1.4	1357	1	RPOB_PSEPU	P19175	pseudomonas
790	6	1.4	1377	1	NEO1_RAT	P97603	rattus norv
791	6	1.4	1391	1	N157_YEAST	P40064	saccharomyc
792	6	1.4	1396	1	ITA2_DROME	P12080	drosophila
793	6	1.4	1433	1	REST_CHICK	O42184	gallus gall
794	6	1.4	1434	1	NOS1_HUMAN	P29475	homo sapien
795	6	1.4	1435	1	NOS1_RABIT	O19132	oryctolagus
796	6	1.4	1456	1	RRPO_PVX	P09395	potato viru
797	6	1.4	1456	1	RRPO_PVXCP	P22591	potato viru
798	6	1.4	1456	1	RRPO_PVXX3	P17779	potato viru
799	6	1.4	1461	1	NEO1_HUMAN	Q92859	homo sapien
800	6	1.4	1476	1	A2M1_MOUSE	P28665	mus musculu
801	6	1.4	1493	1	NEO1_MOUSE	P97798	mus musculu
802	6	1.4	1589	1	PHP_DROME	P39769	drosophila
803	6	1.4	1616	1	GLSF_MAIZE	P23225	zea mays (m
804	6	1.4	1788	1	POLN_SOUV3	Q04544	southampton
805	6	1.4	1938	1	MYSD_CAEEL	P02567	caenorhabdi
806	6	1.4	1966	1	CCAF_HUMAN	O60840	homo sapien
807	6	1.4	1972	1	P531_HUMAN	Q12888	homo sapien
808	6	1.4	2035	1	HFC1_HUMAN	P51610	homo sapien
809	6	1.4	2109	1	RRPL_VSVJH	P13615	vesicular s
810	6	1.4	2109	1	RRPL_VSVJO	P16379	vesicular s
811	6	1.4	2109	1	RRPL_VSVSJ	P03523	vesicular s
812	6	1.4	2142	1	BAT2_HUMAN	P48634	homo sapien
813	6	1.4	2208	1	POLN_MANCV	Q69014	manchester
814	6	1.4	2209	1	Y166_HUMAN	P50748	homo sapien
815	6	1.4	2214	1	PYR1_YEAST	P07259	saccharomyc
816	6	1.4	2244	1	PYR1_SCHPO	Q09794	schizosacch
817	6	1.4	2349	1	TPR_HUMAN	P12270	homo sapien
818	6	1.4	2353	1	CCAH_HUMAN	O95180	homo sapien
819	6	1.4	2376	1	YIM9_YEAST	P40468	saccharomyc
820	6	1.4	2453	1	NCR1_MOUSE	Q60974	mus musculu
821	6	1.4	2483	1	PCX_DROME	P18490	drosophila
822	6	1.4	2524	1	NOTC_XENLA	P21783	xenopus lae
823	6	1.4	2531	1	NTC1_MOUSE	Q01705	mus musculu
824	6	1.4	2531	1	NTC1_RAT	Q07008	rattus norv
825	6	1.4	2813	1	VWF_CANFA	Q28295	canis famil
826	6	1.4	2842	1	APC_RAT	P70478	rattus norv
827	6	1.4	2845	1	APC_MOUSE	Q61315	mus musculu
828	6	1.4	3061	1	POLG_PVYHU	Q02963	p genome po
829	6	1.4	3063	1	POLG_PVYN	P18247	p genome po
830	6	1.4	3068	1	POLG_PEMVC	Q01500	p genome po
831	6	1.4	3110	1	HD_RAT	P51111	rattus norv
832	6	1.4	3119	1	HD_MOUSE	P42859	mus musculu
833	6	1.4	3144	1	HD_HUMAN	P42858	homo sapien
834	6	1.4	3163	1	POLG_TUMVQ	Q02597	t genome po
835	6	1.4	3164	1	POLG_TUMVJ	P89509	t genome po
836	6	1.4	3224	1	RBP2_HUMAN	P49792	homo sapien

837	6	1.4	3358	1	PGCV_MOUSE	Q62059 mus musculu
838	6	1.4	3412	1	POLG_TBEVS	P07720 t genome po
839	6	1.4	4196	1	DYHC_SCHPO	O13290 schizosacch
840	6	1.4	4273	1	PKSM_BACSU	P40872 bacillus su
841	6	1.4	4427	1	PKSL_BACSU	Q05470 bacillus su
842	6	1.4	4447	1	PKSK_BACSU	P40803 bacillus su
843	6	1.4	4590	1	FATH_HUMAN	Q14517 homo sapien

ALIGNMENTS

```
RESULT   1
NU5M_BRALA
ID   NU5M_BRALA     STANDARD;      PRT;   598 AA.
AC   O79422;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN   ND5 OR NAD5 OR NADH5.
OS   Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7740;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98292550; PubMed=9628930;
RA   Spruyt N., Delarbre C., Gachelin G., Laudet V.;
RT   "Complete sequence of the amphioxus (Branchiostoma lanceolatum)
RT   mitochondrial genome: relations to vertebrates.";
RL   Nucleic Acids Res. 26:3279-3285(1998).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y16474; CAA76257.1; -.
DR   InterPro; IPR001516; -.
DR   InterPro; IPR001750; -.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   598 AA;  66097 MW;  EDA986E0B8B1C565 CRC64;


  Query Match             1.8%;  Score 8;  DB 1;  Length 598;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      46 LLLVSAES 53
           ||||||||
Db     122 LLLVSAES 129


RESULT   2
NU5M_BRAFL
ID   NU5M_BRAFL     STANDARD;      PRT;   599 AA.
AC   O47430;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN   ND5 OR NAD5 OR NADH5.
OS   Branchiostoma floridae (Florida lancelet) (Amphioxus).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7739;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99261652; PubMed=10331267;
RA   Boore J.L., Daehler L.L., Brown W.M.;
RT   "Complete sequence, gene arrangement, and genetic code of
RT   mitochondrial DNA of the cephalochordate Branchiostoma floridae
RT   (Amphioxus).";
RL   Mol. Biol. Evol. 16:410-418(1999).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF098298; AAB87999.1; -.
DR   InterPro; IPR001516; -.
DR   InterPro; IPR001750; -.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   599 AA;  66140 MW;  9E04167704FA2990 CRC64;


  Query Match             1.8%;  Score 8;  DB 1;  Length 599;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      46 LLLVSAES 53
           ||||||||
Db     122 LLLVSAES 129


RESULT   3
V120_EBV
ID   V120_EBV       STANDARD;      PRT;  1239 AA.
AC   P03189;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CAPSID ASSEMBLY PROTEIN BOLF1.
GN   BOLF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
CC       EHV-1 23, EBV BOLF1, VZV 21, HVS-1 63, AND HCMV UL47.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; V01555; CAA24841.1; -.
DR   PIR; A03752; QQBE10.
DR   PIR; S32995; S32995.
```

```
KW   Capsid assembly.
SQ   SEQUENCE   1239 AA;. 132748 MW;  6C5DBFC55F2FF729 CRC64;


  Query Match             1.8%;  Score 8;  DB 1;  Length 1239;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     322 RRRLLVPA 329
           ||||||||
Db     874 RRRLLVPA 881


RESULT   4
AKH2_LOCMI
ID   AKH2_LOCMI     STANDARD;      PRT;    61 AA.
AC   P08379; P10617;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ADIPOKINETIC HORMONE II PRECURSOR [CONTAINS: ADIPOKINETIC HORMONE II
DE   (AKH-II); ADIPOKINETIC HORMONE PRECURSOR-RELATED PEPTIDE BETA CHAIN
DE   (APRP-BETA) (6 KDA DIMERIC PEPTIDE B)].
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Corpora cardiaca;
RX   MEDLINE=96032738; PubMed=7559443;
RA   Bogerd J., Kooiman F.P., Pijnenburg M.A.P., Hekking L.H.,
RA   Oudejans R.C., Vander Horst D.J.;
RT   "Molecular cloning of three distinct cDNAs, each encoding a different
RT   adipokinetic hormone precursor, of the migratory locust, Locusta
RT   migratoria. Differential expression of the distinct adipokinetic
RT   hormone precursor genes during flight activity.";
RL   J. Biol. Chem. 270:23038-23043(1995).
RN   [2]
RP   SEQUENCE OF 23-30.
RX   MEDLINE=86050918; PubMed=4063072;
RA   Siegert K., Morgan P., Mordue W.;
RT   "Primary structures of locust adipokinetic hormones II.";
RL   Biol. Chem. Hoppe-Seyler 366:723-727(1985).
RN   [3]
RP   SEQUENCE OF 23-30.
RX   MEDLINE=86130555; PubMed=3947348;
RA   Gaede G., Goldsworthy G.J., Schaffer M.H., Cook J.C.,
RA   Rinehart K.L. Jr.;
RT   "Sequence analyses of adipokinetic hormones II from corpora cardiaca
RT   of Schistocerca nitans, Schistocerca gregaria, and Locusta migratoria
RT   by fast atom bombardment mass spectrometry.";
RL   Biochem. Biophys. Res. Commun. 134:723-730(1986).
RN   [4]
RP   SEQUENCE OF 34-61.
RX   MEDLINE=89276392; PubMed=2731552;
RA   Hietter H., Luu B., Goltzene F., Zachary D., Hoffmann J.A.,
RA   van Dorsselaer A.;
RT   "Isolation and structure of two novel 6-kDa dimeric peptides from the
RT   corpora cardiaca of the insect Locusta migratoria. Molecular mass
RT   determination by mass spectrometry.";
RL   Eur. J. Biochem. 182:77-84(1989).
CC   -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC       CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC       DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC       MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC   -!- SUBUNIT: ADIPOKINETIC HORMONE PRECURSOR-RELATED PEPTIDE (APRP)
CC       CAN FORM THREE TYPE OF DISULFIDE-BOND DIMERS: P1 (ALPHA-ALPHA),
CC       P2 (ALPHA-BETA), AND P3 (BETA-BETA).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X86800; CAA60495.1; -.
DR   PIR; B25204; AKLQL2.
DR   PIR; B24241; B24241.
DR   InterPro; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Flight; Cleavage on pair of basic residues;
KW   Signal.
FT   SIGNAL        1     22
FT   CHAIN        23     61       ADIPOKINETIC HORMONE PRO-HORMONE CHAIN B.
FT   PEPTIDE      23     30       ADIPOKINETIC HORMONE II.
FT   PEPTIDE      34     61       ADIPOKINETIC HORMONE PRECURSOR-RELATED
FT                                PEPTIDE BETA CHAIN (6 KDA DIMERIC PEPTIDE
FT                                B).
FT   MOD_RES      23     23       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      30     30       AMIDATION (G-31 PROVIDE AMIDE GROUP).
FT   DISULFID     59     59       INTERCHAIN.
SQ   SEQUENCE   61 AA;  6588 MW;  3436B59C3712A046 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 TLVLVVA 43
           |||||||
Db       8 TLVLVVA 14


RESULT   5
RL22_PRUAP
ID   RL22_PRUAP     STANDARD;      PRT;    79 AA.
AC   Q44161;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L22 (FRAGMENT).
GN   RPLV OR RPL22.
OS   Prunus armeniaca phytoplasma.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Acholeplasmataceae; Phytoplasma.
OX   NCBI_TaxID=36589;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94350802; PubMed=8071198;
RA   Gundersen D.E., Lee I.M., Rehner S.A., Davis R.E., Kingsbury D.T.;
RT   "Phylogeny of mycoplasmalike organisms (phytoplasmas): a basis for
RT   their classification.";
RL   J. Bacteriol. 176:5244-5254(1994).
CC   -!- FUNCTION: THIS PROTEIN BINDS SPECIFICALLY TO 23S RRNA; ITS BINDING
CC       IS STIMULATED BY OTHER RIBOSOMAL PROTEINS, E.G., L4, L17, AND L20.
CC       IT IS IMPORTANT DURING THE EARLY STAGES OF 50S RECONSTITUTION
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L22P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L26994; AAA83935.1; -.
DR   HSSP; P48286; 1BXE.
DR   Pfam; PF00237; Ribosomal_L22; 1.
DR   PROSITE; PS00464; RIBOSOMAL_L22; PARTIAL.
```

```
KW   Ribosomal protein; rRNA-binding.
FT   NON_TER      79     79
SQ   SEQUENCE   79 AA;  8687 MW;  0DB55502D51EA166 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 79;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     367 EIKVAKA 373
           |||||||
Db      28 EIKVAKA 34


RESULT   6
SINR_BACSU
ID   SINR_BACSU     STANDARD;      PRT;   111 AA.
AC   P06533;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SINR PROTEIN.
GN   SINR OR SIN OR FLAD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87056981; PubMed=3096962;
RA   Gaur N.K., Dubnau E., Smith I.;
RT   "Characterization of a cloned Bacillus subtilis gene that inhibits
RT   sporulation in multiple copies.";
RL   J. Bacteriol. 168:860-869(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RA   Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA   Sato T., Takeuchi M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   FUNCTION.
RX   MEDLINE=91100356; PubMed=1898931;
RA   Gaur N.K., Oppenheim J., Smith I.;
RT   "The Bacillus subtilis sin gene, a regulator of alternate
RT   developmental processes, codes for a DNA-binding protein.";
RL   J. Bacteriol. 173:678-686(1991).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=95370138; PubMed=7642487;
RA   Mandic-Mulec I., Doukhan L., Smith I.;
RT   "The Bacillus subtilis SinR protein is a repressor of the key
RT   sporulation gene spo0A.";
RL   J. Bacteriol. 177:4619-4627(1995).
CC   -!- FUNCTION: NEGATIVE AS WELL AS POSITIVE REGULATOR OF ALTERNATE
CC       DEVELOPMENTAL PROCESSES THAT ARE INDUCED AT THE END OF VEGETATIVE
CC       GROWTH IN RESPONSE TO NUTRIENT DEPLETION. BINDS TO THE ALKALINE
CC       PROTEASE (APRE) GENE AT TWO SITES. ALSO ACTS AS A REPRESSOR OF THE
CC       SPO0A GENE.
CC   -!- SUBUNIT: HOMOTETRAMER. ALSO ASSOCIATES WITH SINI.
CC   -!- SIMILARITY: BELONGS TO THE PBSX(XRE) FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14112; AAA22757.1; -.
DR   EMBL; D84432; BAA12542.1; -.
DR   EMBL; Z99116; CAB14392.1; -.
DR   PIR; B25159; B25159.
DR   SubtiList; BG10754; sinR.
DR   InterPro; IPR001387; -.
DR   Pfam; PF01381; HTH_3; 1.
KW   Transcription regulation; DNA-binding; Activator; Repressor;
KW   Sporulation.
FT   DNA_BIND     17     36       H-T-H MOTIF (POTENTIAL).
SQ   SEQUENCE   111 AA;  12989 MW;  B6FC6463FF4B52C6 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     398 SVHTLLD 404
           |||||||
Db      57 SVHTLLD 63


RESULT   7
GP38_CANFA
ID   GP38_CANFA     STANDARD;      PRT;   169 AA.
AC   Q95152;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   GLYCOPROTEIN 38 PRECURSOR (GP38) (MUCIN-TYPE MEMBRANE PROTEIN GP40).
GN   GP38 OR GP40.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Kidney;
RX   MEDLINE=97479204; PubMed=9337856;
RA   Zimmer G., Lottspeich F., Maisner A., Klenk H.D., Herrler G.;
RT   "Molecular characterization of gp40, a mucin-type glycoprotein from
RT   the apical plasma membrane of Madin-Darby canine kidney cells (type
RT   I).";
RL   Biochem. J. 326:99-108(1997).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z81018; CAB02675.1; -.
KW   Signal; Transmembrane; Glycoprotein.
FT   SIGNAL       1     16       POTENTIAL.
FT   CHAIN       17    169       GLYCOPROTEIN 38.
FT   DOMAIN      17    138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    139   159       POTENTIAL.
FT   DOMAIN      160   169       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   169 AA;  17574 MW;  BB2D48FE55C56DBC CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 169;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     265 PGAEDNV 271
           |||||||
Db      51 PGAEDNV 57


RESULT   8
```

```
NUC_SHIFL
ID   NUC_SHIFL      STANDARD;      PRT;   174 AA.
AC   P29769;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   MICROCOCCAL NUCLEASE PRECURSOR (EC 3.1.31.1).
GN   NUC.
OS   Shigella flexneri.
OG   Plasmid IncW pSa.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92219997; PubMed=1560781;
RA   Close S.M., Kado C.I.;
RT   "A gene near the plasmid pSa origin of replication encodes a
RT   nuclease.";
RL   Mol. Microbiol. 6:521-527(1992).
CC   -!- SIMILARITY: BELONGS TO THE THERMONUCLEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30471; AAA75246.1; -.
DR   PIR; S20541; NCEBPX.
DR   HSSP; P00644; 1SNQ.
DR   InterPro; IPR000080; -.
DR   InterPro; IPR002071; -.
DR   Pfam; PF00565; SNase; 1.
DR   PROSITE; PS01123; TNASE_1; 1.
DR   PROSITE; PS01284; TNASE_2; 1.
KW   Hydrolase; Nuclease; Endonuclease; Signal; Plasmid.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    174       MICROCOCCAL NUCLEASE.
FT   ACT_SITE     52     52       BY SIMILARITY.
FT   ACT_SITE     60     60       BY SIMILARITY.
FT   ACT_SITE     94     94       BY SIMILARITY.
SQ   SEQUENCE   174 AA;  19689 MW;  9B2B37729E4F30D7 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 174;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     221 AAVVLIV 227
           |||||||
Db      13 AAVVLIV 19


RESULT   9
RIMM_MYCLE
ID   RIMM_MYCLE     STANDARD;      PRT;   179 AA.
AC   O33016;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROBABLE 16S RRNA PROCESSING PROTEIN RIMM.
GN   RIMM OR MLCB250.34.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K.J., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL FOR EFFICIENT PROCESSING OF 16S RRNA. PROBABLY
CC       PART OF THE 30S SUBUNIT PRIOR TO OR DURING THE FINAL STEP IN THE
CC       PROCESSING OF 16S FREE 30S RIBOSOMAL SUBUNITS. IT COULD BE SOME
CC       ACCESSORY PROTEIN NEEDED FOR EFFICIENT ASSEMBLY OF THE 30S
CC       SUBUNIT. RIMM IS NEEDED IN A STEP PRIOR TO RBFA DURING THE
CC       MATURATION OF 16S RRNA. HAS AFFINITY FOR FREE RIBOSOMAL 30S
CC       SUBUNITS BUT NOT FOR 70S RIBOSOMES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE RIMM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z97369; CAB10628.1; -.
DR   InterPro; IPR002676; -.
DR   Pfam; PF01782; RimM; 1.
KW   rRNA processing.
SQ   SEQUENCE   179 AA;  19295 MW;  F95FFD5C3735930E CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 179;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     118 RCDSGEV 124
           |||||||
Db     143 RCDSGEV 149


RESULT  10
YDEQ_BACSU
ID   YDEQ_BACSU     STANDARD;      PRT;   197 AA.
AC   P96674;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PUTATIVE NAD(P)H OXIDOREDUCTASE YDEQ (EC 1.6.99.-).
GN   YDEQ.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kasahara Y., Nakai S., Lee S., Sadaie Y., Ogasawara N.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE NAD(P)H DEHYDROGENASE (QUINONE) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB001488; BAA19364.1; -.
DR   EMBL; Z99106; CAB12337.1; -.
DR   SubtiList; BG12144; ydeQ.
KW   Hypothetical protein; Oxidoreductase.
SQ   SEQUENCE   197 AA;  22831 MW;  B6BFE72C4AFFBFCE CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 197;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      36 KTLVLVV 42
           |||||||
Db       5 KTLVLVV 11


RESULT  11
YO27_CAEEL
ID   YO27_CAEEL     STANDARD;      PRT;   243 AA.
AC   P34677;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   HYPOTHETICAL 26.5 KDA PROTEIN ZK688.7 IN CHROMOSOME III.
GN   ZK688.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: WITH C13G5.2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L16621; AAA28225.1; -.
DR   PIR; S44914; S44914.
DR   WormPep; ZK688.7; CE00465.
DR   InterPro; IPR000535; -.
DR   Pfam; PF00635; MSP_domain; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   243 AA;  26480 MW;  27243B16EA74B1C7 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     131 TTRNTVC 137
           |||||||
Db     142 TTRNTVC 148


RESULT  12
MTDH_UROFA
ID   MTDH_UROFA     STANDARD;      PRT;   256 AA.
AC   O00058;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PROBABLE NADP-DEPENDENT MANNITOL DEHYDROGENASE (EC 1.1.1.138) (MTDH)
DE   (MANNITOL 2-DEHYDROGENASE [NADP+]) (PLANTA-INDUCED RUST PROTEIN 8).
GN   PIG8.
OS   Uromyces fabae.
OC   Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Uredinales;
OC   Pucciniaceae; Uromyces.
OX   NCBI_TaxID=55588;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=I2; TISSUE=Haustorium;
RX   MEDLINE=97294919; PubMed=9150592;
RA   Hahn M., Mendgen K.;
RT   "Characterization of in planta-induced rust genes isolated from a
RT   haustorium-specific cDNA library.";
RL   Mol. Plant Microbe Interact. 10:427-437(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN HEXITOL OR PENTITOL METABOLISM.
CC   -!- CATALYTIC ACTIVITY: D-MANNITOL + NADP(+) = D-FRUCTOSE + NADPH.
CC   -!- DEVELOPMENTAL STAGE: HAUSTORIA AND RUST-INFECTED LEAVES.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U81790; AAB39878.1; -.
DR   InterPro; IPR002198; -.
DR   InterPro; IPR002347; -.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00678; adh_short_C2; 1.
DR   PRINTS; PR00080; SDRFAMILY.
DR   PRINTS; PR00081; GDHRDH.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Oxidoreductase; NADP.
FT   NP_BIND      15     37       NADP (BY SIMILARITY).
FT   ACT_SITE    163    163       BY SIMILARITY.
SQ   SEQUENCE   256 AA;  27842 MW;  218677F0010364D5 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     300 VNMLSPG 306
           |||||||
Db     188 VNMLSPG 194


RESULT  13
ARCC_ECOLI
ID   ARCC_ECOLI     STANDARD;      PRT;   297 AA.
AC   P37306; P77419;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CARBAMATE KINASE (EC 2.7.2.2).
GN   ARCC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 203-297 FROM N.A.
RX   MEDLINE=89123018; PubMed=2644189;
RA   Watanabe W., Sampei G., Aiba A., Mizobuchi K.;
RT   "Identification and sequence analysis of Escherichia coli purE and
RT   purK genes encoding 5'-phosphoribosyl-5-amino-4-imidazole carboxylase
RT   for de novo purine biosynthesis.";
RL   J. Bacteriol. 171:198-204(1989).
RN   [4]
RP   SEQUENCE OF 203-297 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=89123019; PubMed=2464576;
RA   Tiedeman A.A., Keyhani J., Kamholz J., Daum H.A. III, Gots J.S.,
RA   Smith J.M.;
RT   "Nucleotide sequence analysis of the purEK operon encoding 5'-
RT   phosphoribosyl-5-aminoimidazole carboxylase of Escherichia coli
RT   K-12.";
RL   J. Bacteriol. 171:205-212(1989).
RN   [5]
RP   IDENTIFICATION.
RX   MEDLINE=95026987; PubMed=7940673;
RA   Borodovsky M., Koonin E.V., Rudd K.E.;
RT   "New genes in old sequence: a strategy for finding genes in the
RT   bacterial genome.";
RL   Trends Biochem. Sci. 19:309-313(1994).
RN   [6]
RP   IDENTIFICATION.
RX   MEDLINE=95075659; PubMed=7984428;
RA   Borodovsky M., Rudd K.E., Koonin E.V.;
RT   "Intrinsic and extrinsic approaches for detecting genes in a
RT   bacterial genome.";
RL   Nucleic Acids Res. 22:4756-4767(1994).
RN   [7]
RP   IDENTIFICATION.
RX   MEDLINE=95004589; PubMed=7920643;
RA   Robison K., Gilbert W., Church G.M.;
RT   "Large scale bacterial gene discovery by similarity search.";
RL   Nat. Genet. 7:205-214(1994).
CC   -!- CATALYTIC ACTIVITY: ATP + NH(3) + CO(2) = ADP + CARBAMOYL
CC       PHOSPHATE.
CC   -!- PATHWAY: THIRD STEP IN ARGININE DEGRADATION VIA ARGININE
CC       DEIMINASE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CARBAMATE KINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000158; AAC73623.1; -.
DR   EMBL; U82664; AAB40273.1; -.
DR   EMBL; M19657; -; NOT_ANNOTATED_CDS.
DR   EMBL; X12982; -; NOT_ANNOTATED_CDS.
DR   EcoGene;.EG12384; ybcF.
KW   Transferase; Kinase; Arginine metabolism.
FT   CONFLICT    279    297       ALSRIEETLAGEAGTCISL -> RYRELKRRWRAKRGPVFR
FT                                CSRRH (IN REF. 4).
SQ   SEQUENCE   297 AA;  31644 MW;  0169371E119AE832 CRC64;

  Query Match             1.6%;  Score 7;  DB 1;  Length 297;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     247 ICSGGGG 253
           |||||||
Db     180 ICSGGGG 186


RESULT  14
METE_METTH
ID   METE_METTH     STANDARD;      PRT;   308 AA.
AC   O26869;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PROBABLE METHYLCOBALAMIN:HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.-)
DE   (METHIONINE SYNTHASE).
GN   METE OR MTH775.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM
CC       METHYLCOBALAMIN AND METHYLCOBINAMIDE TO HOMOCYSTEINE RESULTING IN
CC       METHIONINE FORMATION (BY SIMILARITY).
CC   -!- COFACTOR: ZINC; BINDS ONE MOLE PER SUBUNIT (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC   -!- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC       SYNTHASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000856; AAB85278.1; -.
DR   InterPro; IPR002629; -.
DR   Pfam; PF01717; Methionine_synt; 1.
KW   Transferase; Methyltransferase; Methionine biosynthesis; Zinc.
FT   METAL       201    201       ZINC (BY SIMILARITY).
FT   METAL       203    203       ZINC (BY SIMILARITY).
FT   METAL       285    285       ZINC (BY SIMILARITY).
SQ   SEQUENCE   308 AA;  33704 MW;  0157F63D21C80F4F CRC64;

  Query Match             1.6%;  Score 7;  DB 1;  Length 308;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     407 ETLGERL 413
           |||||||
Db      17 ETLGERL 23


RESULT  15
MER4_STRLI
```

```
ID   MER4_STRLI     STANDARD;      PRT;   319 AA.
AC   P30344;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MERCURY RESISTANCE PROBABLE HG TRANSPORT PROTEIN.
OS   Streptomyces lividans.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1916;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=66 / 1326;
RX   MEDLINE=93156687; PubMed=1494353;
RA   Sedlmeier R., Altenbuchner J.;
RT   "Cloning and DNA sequence analysis of the mercury resistance genes of
RT   Streptomyces lividans.";
RL   Mol. Gen. Genet. 236:76-85(1992).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65467; CAA46465.1; -.
DR   PIR; S23613; S23613.
DR   PIR; S30173; S30173.
DR   HSSP; P04002; 1WFA.
KW   Mercury; Mercuric resistance.
FT   METAL       298    298       HG(2+) (POTENTIAL).
FT   METAL       299    299       HG(2+) (POTENTIAL).
FT   METAL       318    318       HG(2+) (POTENTIAL).
FT   METAL       319    319       HG(2+) (POTENTIAL).
SQ   SEQUENCE   319 AA;  31552 MW;  031495174FD15A14 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 319;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      43 AAVLLLV 49
           |||||||
Db     174 AAVLLLV 180


RESULT  16
HOLB_PSEAE
ID   HOLB_PSEAE     STANDARD;      PRT;   328 AA.
AC   P52024;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DNA POLYMERASE III, DELTA' SUBUNIT (EC 2.7.7.7).
GN   HOLB OR PA2961.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=96125227; PubMed=8550441;
RA   Alm R.A., Bodero A.J., Free P.D., Mattick J.S.;
RT   "Identification of a novel gene, pilZ, essential for type 4 fimbrial
RT   biogenesis in Pseudomonas aeruginosa.";
RL   J. Bacteriol. 178:46-53(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC       N PYROPHOSPHATE + DNA(N).
CC   -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC       CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC       DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC       THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC       AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L42622; AAA93518.1; -.
DR   EMBL; AE004722; AAG06349.1; -.
DR   HSSP; P28631; 1A5T.
KW   Transferase; DNA-directed DNA polymerase; DNA replication.
FT   CONFLICT     16     30       GGRAQHAHAYLLYGP -> RPRPARPRLSALAS (IN
FT                                REF. 1).
FT   CONFLICT     48     52       CQRPA -> TRAC (IN REF. 1).
FT   CONFLICT    176    177       RA -> PG (IN REF. 1).
FT   CONFLICT    208    209       VR -> AG (IN REF. 1).
FT   CONFLICT    238    251       SVPLPLLFDWFCDW -> RVPIAIALRLVLRL (IN REF.
FT                                1).
FT   CONFLICT    285    286       QA -> HR (IN REF. 1).
FT   CONFLICT    322    322       A -> G (IN REF. 1).
SQ   SEQUENCE   328 AA;  35746 MW;  55DAA4F223C1464D CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     311 LLEPAEA 317
           |||||||
Db     111 LLEPAEA 117


RESULT  17
IF2B_RABIT
ID   IF2B_RABIT     STANDARD;      PRT;   333 AA.
AC   P41035;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA).
GN   EIF2S2 OR EIF2B.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE;
RX   MEDLINE=94032484; PubMed=8218412;
RA   Price N.T., Hall L., Proud C.G.;
RT   "Cloning of cDNA for the beta-subunit of rabbit translation initiation
```

```
RT   factor-2 using PCR.";
RL   Biochim. Biophys. Acta 1216:170-172(1993).
RN   [2]
RP   SEQUENCE OF 25-39; 122-136 AND 195-218.
RX   MEDLINE=92002150; PubMed=1911855;
RA   Bommer U.A., Kraft R., Kurzchalia T.V., Price N.T., Proud C.G.;
RT   "Amino acid sequence analysis of the beta- and gamma-subunits of
RT   eukaryotic initiation factor eIF-2. Identification of regions
RT   interacting with GTP.";
RL   Biochim. Biophys. Acta 1079:308-315(1991).
CC   -!- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS
CC       BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS
CC       COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING
CC       TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL
CC       SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY
CC       HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP
CC       BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER
CC       ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP
CC       BY WAY OF A REACTION CATALYZED BY EIF-2B.
CC   -!- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
CC       CHAIN.
CC   -!- SIMILARITY: BELONGS TO THE EIF2B / EIF5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73836; CAA52058.1; -.
DR   PIR; S38536; S38536.
DR   PIR; S17871; S17871.
DR   InterPro; IPR002735; -.
DR   Pfam; PF01873; eIF5_eIF2B; 1.
KW   Initiation factor; Protein biosynthesis; Zinc-finger.
FT   DOMAIN       14     21       POLY-LYS (BASIC).
FT   DOMAIN       79     87       POLY-LYS (BASIC).
FT   DOMAIN      124    129       POLY-LYS (BASIC).
FT   ZN_FING     281    305       C4-TYPE (POTENTIAL).
SQ   SEQUENCE   333 AA;  38326 MW;  3FFDBAB92DFC1465 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 333;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     291 VQEPAEP 297
           |||||||
Db     107 VQEPAEP 113


RESULT  18
CAHC_HUMAN
ID   CAHC_HUMAN     STANDARD;      PRT;   354 AA.
AC   O43570;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CARBONIC ANHYDRASE XII PRECURSOR (EC 4.2.1.1) (CARBONATE DEHYDRATASE
DE   XII) (CA-XII) (TUMOR ANTIGEN HOM-RCC-3.1.3).
GN   CA12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98301622; PubMed=9636197;
RA   Tureci O., Sahin U., Vollmar E., Siemer S., Gottert E., Seitz G.,
RA   Parkkila A.K., Shah G.N., Grubb J.H., Pfreundschuh M., Sly W.S.;
RT   "Human carbonic anhydrase XII: cDNA cloning, expression, and
RT   chromosomal localization of a carbonic anhydrase gene that is
RT   overexpressed in some renal cell cancers.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:7608-7613(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98445416; PubMed=9770531;
RA   Ivanov S.V., Kuzmin I., Wei M.-H., Pack S., Geil L., Johnson B.E.,
RA   Stanbridge E.J., Lerman M.I.;
RT   "Down-regulation of transmembrane carbonic anhydrases in renal cell
RT   carcinoma cell lines by wild-type von Hippel-Lindau transgenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12596-12601(1998).
CC   -!- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC   -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN COLON, KIDNEY, AND
CC       PROSTATE AND MODERATELY EXPRESSED IN PANCREAS, OVARY, AND TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF051882; AAC39789.1; -.
DR   EMBL; AF037335; AAC63952.1; -.
DR   MIM; 603263; -.
DR   InterPro; IPR001148; -.
DR   Pfam; PF00194; carb_anhydrase; 1.
DR   PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW   Lyase; Zinc; Transmembrane; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    354       CARBONIC ANHYDRASE XII.
FT   DOMAIN       25    301       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    302    322       POTENTIAL.
FT   DOMAIN      323    354       CYTOPLASMIC (POTENTIAL).
FT   METAL       119    119       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       121    121       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       145    145       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    162    162       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   354 AA;  39451 MW;  9016216BF2CA6C0C CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      43 AAVLLLV 49
           |||||||
Db       9 AAVLLLV 15


RESULT  19
SUR6_HUMAN
ID   SUR6_HUMAN     STANDARD;      PRT;   361 AA.
AC   O75683; Q9UK24;
DT   15-JUL-1999 (Rel. 38, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   SURFEIT LOCUS PROTEIN 6.
GN   SURF6 OR SURF-6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Magoulas C., Fried M.;
```

```
RT   "Isolation and genomic analysis of the human Surf-6 gene: a member of
RT   the surfeit locus.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-31 FROM N.A.
RA   Angiolillo A., Russo G., Porcellini A., Smaldone S., Scognamiglio B.,
RA   D'Alessandro F., Pietropaolo C.T.;
RT   "The human homologue of the mouse Surf-5 gene encodes multiple
RT   alternatively spliced transcripts.";
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN A NUCLEOLAR FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
CC   -!- SIMILARITY: BELONGS TO THE SURF6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF186772; AAD56587.1; -.
DR   EMBL; AJ224639; CAA12054.1; -.
DR   MIM; 185642; -.
KW   Nuclear protein.
FT   DOMAIN       45     49       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      154    158       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   CONFLICT      6      6       A -> T (IN REF. 2).
SQ   SEQUENCE   361 AA;  41449 MW;  5B9639916E882791 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 PAASGAR 15
           |||||||
Db      75 PAASGAR 81


RESULT  20
METX_CORGL
ID   METX_CORGL     STANDARD;      PRT;   379 AA.
AC   O68640;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31) (HOMOSERINE O-TRANS-
DE   ACETYLASE) (HOMOSERINE TRANSACETYLASE) (HTA).
GN   METX OR META.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC   Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=98330894; PubMed=9666465;
RA   Park S.-D., Lee J.-Y., Kim Y., Kim J.-H., Lee H.-S.;
RT   "Isolation and analysis of metA, a methionine biosynthetic gene
RT   encoding homoserine acetyltransferase in corynebacterium glutamicum.";
RL   Mol. Cells 8:286-294(1998).
CC   -!- CATALYTIC ACTIVITY: ACETYL-COA + L-HOMOSERINE = COA + O-ACETYL-L-
CC       HOMOSERINE.
CC   -!- PATHWAY: BIOSYNTHESIS OF METHIONINE; HTA VARIANT; FIRST STEP.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE AB HYDROLASE FAMILY; HTA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF052652; AAC06035.1; -.
DR   InterPro; IPR000073; -.
DR   Pfam; PF00561; abhydrolase; 1.
KW   Methionine biosynthesis; Transferase; Acyltransferase.
FT   ACT_SITE    145    145       POTENTIAL.
FT   ACT_SITE    343    343       POTENTIAL.
SQ   SEQUENCE   379 AA;  41390 MW;  4ABB500D19E649F1 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 379;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     219 TVAAVVL 225
           |||||||
Db     135 TVAAVVL 141


RESULT  21
CYCH_RHIME
ID   CYCH_RHIME     STANDARD;      PRT;   381 AA.
AC   P45400;
DT   01-NOV-1995 (Rel. 32, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCH.
GN   CYCH.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK631;
RX   MEDLINE=95231514; PubMed=7715602;
RA   Kereszt A., Slaska-Kiss K., Putnoky P., Banfalvi Z., Kondorosi A.;
RT   "The cycHJKL genes of Rhizobium meliloti involved in cytochrome c
RT   biogenesis are required for 'respiratory' nitrate reduction ex planta
RT   and for nitrogen fixation during symbiosis.";
RL   Mol. Gen. Genet. 247:39-47(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA   Boistard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,
RA   Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA   Vandenbol M., Puehler A., Becker A., Weidner S.;
RL   Submitted (MAR-2000) to the SWISS-PROT data bank.
CC   -!- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC       POSSIBLE SUBUNIT OF A HEME LYASE.
CC   -!- SUBCELLULAR LOCATION: MOSTLY PERIPLASMIC; ANCHORED IN THE INNER
CC       MEMBRANE (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82560; CAA57904.1; -.
KW   Cytochrome c-type biogenesis; Transmembrane; Inner membrane.
FT   TRANSMEM      2     22       POTENTIAL.
FT   TRANSMEM     95    115       POTENTIAL.
FT   DOMAIN      116    381       PERIPLASMIC (POTENTIAL).
FT   CONFLICT     74     74       S -> A (IN REF. 1).
FT   CONFLICT    226    226       R -> P (IN REF. 1).
FT   CONFLICT    240    240       K -> R (IN REF. 1).
```

```
FT   CONFLICT    271    272       EL -> DV (IN REF. 1).
FT   CONFLICT    280    281       MISSING (IN REF. 1).
FT   CONFLICT    286    286       D -> N (IN REF. 1).
FT   CONFLICT    329    329       M -> I (IN REF. 1).
SQ   SEQUENCE   381 AA;  40321 MW;  DFFB39391554033E CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 VAAVLLL 48
           |||||||
Db      14 VAAVLLL 20


RESULT  22
NR23_MOUSE
ID   NR23_MOUSE     STANDARD;      PRT;   395 AA.
AC   Q9QXZ7;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PHOTORECEPTOR-SPECIFIC NUCLEAR RECEPTOR (RETINA-SPECIFIC NUCLEAR
DE   RECEPTOR).
GN   NR2E3 OR PNR OR RNR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Retina;
RX   MEDLINE=20079622; PubMed=10611353;
RA   Chen F., Figueroa D.J., Marmorstein A.D., Zhang Q., Petrukhin K.,
RA   Caskey C.T., Austin C.P.;
RT   "Retina-specific nuclear receptor: A potential regulator of cellular
RT   retinaldehyde-binding protein expressed in retinal pigment epithelium
RT   and Muller glial cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:15149-15154(1999).
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20266403; PubMed=10805811;
RA   Akhmedov N.B., Piriev N.I., Chang B., Rapoport A.L., Hawes N.L.,
RA   Nishina P.M., Nusinowitz S., Heckenlively J.R., Roderick T.H.,
RA   Kozak C.A., Danciger M., Davisson M.T., Farber D.B.;
RT   "A deletion in a photoreceptor-specific nuclear receptor mRNA causes
RT   retinal degeneration in the rd7 mouse.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:5551-5556(2000).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT MAY HAVE A ROLE IN THE
CC       REGULATION OF SIGNALING PATHWAYS INTRINSIC TO THE PHOTORECEPTOR
CC       CELL FUNCTION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: RETINA-SPECIFIC.
CC   -!- DISEASE: DEFECTS IN THIS PROTEIN ARE THE CAUSE OF RETINAL
CC       DEGENERATION.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF148129; AAF22228.1; -.
DR   EMBL; AF204053; AAF69682.1; -.
DR   MGD; MGI:1346317; Nr2e3.
DR   InterPro; IPR000003; -.
DR   InterPro; IPR000324; -.
DR   InterPro; IPR000536; -.
DR   InterPro; IPR001628; -.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   PRINTS; PR00350; VITAMINDR.
DR   PRINTS; PR00545; RETINOIDXR.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger.
FT   DNA_BIND     40    107       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      40     60       C4-TYPE.
FT   ZN_FING      76    102       C4-TYPE.
SQ   SEQUENCE   395 AA;  43176 MW;  D8E391CF24919A68 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 395;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     205 PASPCSL 211
           |||||||
Db     198 PASPCSL 204


RESULT  23
LEF1_MOUSE
ID   LEF1_MOUSE     STANDARD;      PRT;   397 AA.
AC   P27782;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   LYMPHOID ENHANCER BINDING FACTOR 1 (LEF-1).
GN   LEF1 OR LEF-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6 X DBA;
RX   MEDLINE=91224496; PubMed=1827423;
RA   Travis A., Amsterdam A., Belanger C., Grosschedl R.;
RT   "LEF-1, a gene encoding a lymphoid-specific protein with an HMG
RT   domain, regulates T-cell receptor alpha enhancer function.";
RL   Genes Dev. 5:880-894(1991).
RN   [2]
RP   STRUCTURE BY NMR OF 296-380.
RX   MEDLINE=95379976; PubMed=7651541;
RA   Love J.J., Li X., Case D.A., Glese K., Grosschedl P., Wright P.E.;
RT   "Structural basis for DNA bending by the architectural transcription
RT   factor LEF-1.";
RL   Nature 376:791-795(1995).
RN   [3]
RP   STRUCTURE BY NMR OF 296-380.
RA   Li X., Love J.J., Case D.A., Wright P.E.;
RT   "High resolution NMR structure of the LEF-1 HMG domain complexed with
RT   its cognate DNA.";
RL   Submitted (OCT-1998) to the PDB data bank.
CC   -!- FUNCTION: REGULATES T-CELL RECEPTOR ALPHA ENHANCER FUNCTION.
CC       BINDS DNA IN A SEQUENCE-SPECIFIC MANNER.
CC   -!- SUBUNIT: INTERACTS WITH BETA-CATENIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: LYMPHOCYTES.
CC   -!- DOMAIN: PROLINE-RICH AND ACIDIC REGIONS ARE IMPLICATED IN THE
CC       ACTIVATION FUNCTIONS OF RNA POLYMERASE II TRANSCRIPTION FACTORS.
CC   -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58636; CAA41493.1; -.
DR   PIR; A39565; A39565.
DR   PDB; 2LEF; 21-OCT-98.
DR   TRANSFAC; T00930; -.
DR   MGD; MGI:96770; Lef1.
DR   InterPro; IPR000910; -.
DR   Pfam; PF00505; HMG_box; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation; 3D-structure.
FT   DOMAIN        6     11       POLY-GLY.
FT   DOMAIN       12     50       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       75    271       PRO-RICH.
FT   DNA_BIND    297    365       HMG BOX.
FT   DOMAIN      372    377       POLY-LYS.
SQ   SEQUENCE   397 AA;  44058 MW;  0CA01B6C528FD7FA CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     250 GGGGDPE 256
           |||||||
Db       8 GGGGDPE 14


RESULT  24
LEF1_RAT
ID   LEF1_RAT       STANDARD;      PRT;   397 AA.
AC   Q9QXN1;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   LYMPHOID ENHANCER BINDING FACTOR 1 (LEF-1).
GN   LEF1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Kobielak K., Kobielak A., Trzeciak W.H.;
RT   "Clonig of lymphoid enhancer binding factor-1 (LEF-1) from rat kidney:
RT   homology to the murine sequence.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REGULATES T-CELL RECEPTOR ALPHA ENHANCER FUNCTION.
CC       BINDS DNA IN A SEQUENCE-SPECIFIC MANNER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF198533; AAF15601.1; -.
KW   DNA-binding; Nuclear protein; Transcription regulation.
FT   DOMAIN        6     11       POLY-GLY.
FT   DOMAIN       12     50       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       75    271       PRO-RICH.
FT   DNA_BIND    297    365       HMG BOX.
FT   DOMAIN      372    377       POLY-LYS.
SQ   SEQUENCE   397 AA;  44023 MW;  D6BEFF805CB526EA CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     250 GGGGDPE 256
           |||||||
Db       8 GGGGDPE 14


RESULT  25
LEF1_HUMAN
ID   LEF1_HUMAN     STANDARD;      PRT;   399 AA.
AC   Q9UJU2;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   LYMPHOID ENHANCER BINDING FACTOR 1 (LEF-1) (T CELL-SPECIFIC
DE   TRANSCRIPTION FACTOR 1-ALPHA) (TCF1-ALPHA).
GN   LEF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kobielak A., Kobielak K., Trzeciak W.H.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REGULATES T-CELL RECEPTOR ALPHA ENHANCER FUNCTION.
CC       BINDS DNA IN A SEQUENCE-SPECIFIC MANNER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF198532; AAF13268.1; -.
DR   MIM; 153245; -.
KW   DNA-binding; Nuclear protein; Transcription regulation.
FT   DOMAIN        6     13       POLY-GLY.
FT   DOMAIN       14     52       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       77    273       PRO-RICH.
FT   DNA_BIND    299    367       HMG BOX.
FT   DOMAIN      374    379       POLY-LYS.
SQ   SEQUENCE   399 AA;  44200 MW;  D480D440698EEFE3 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 399;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     250 GGGGDPE 256
           |||||||
Db      10 GGGGDPE 16


RESULT  26
Y413_ARATH
ID   Y413_ARATH     STANDARD;      PRT;   435 AA.
AC   O04656;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 48.8 KDA PROTEIN M021B04.13.
GN   A_TM021B04.13.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Dante M., Wamsley P., Gibson A.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF007271; AAB61074.1; -.
DR   InterPro; IPR002687; -.
DR   Pfam; PF01798; Nop; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   435 AA;  48818 MW;  BE502C8C244AF085 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 435;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    174 SDIECVH 180
          |||||||
Db     45 SDIECVH 51


RESULT  27
SYD_METTH
ID   SYD_METTH      STANDARD;      PRT;   437 AA.
AC   O26328;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12) (ASPARTATE--TRNA LIGASE)
DE   (ASPRS).
GN   ASPS OR MTH226.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +
CC       PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000809; AAB84732.1; -.
DR   HSSP; P04802; 1ASY.
DR   InterPro; IPR002106; -.
DR   InterPro; IPR002309; -.
DR   InterPro; IPR002312; -.
DR   Pfam; PF00152; tRNA-synt_2; 1.
DR   PRINTS; PR01042; TRNASYNTHASP.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ   SEQUENCE   437 AA;  49838 MW;  128C9BE08FBFA6D3 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 437;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    404 DALETLG 410
          |||||||
Db    268 DALETLG 274


RESULT  28
RECN_HELPJ
ID   RECN_HELPJ      STANDARD;      PRT;   522 AA.
AC   Q9ZJ80;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N).
GN   RECN OR JHP1434.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
CC       DNA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RECN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001566; AAD07019.1; -.
DR   InterPro; IPR001617; -.
KW   DNA repair; ATP-binding.
FT   NP_BIND      36     43       ATP (POTENTIAL).
SQ   SEQUENCE   522 AA;  59349 MW;  000D7EE3C0632EA5 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 522;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    402 LLDALET 408
          |||||||
Db    156 LLDALET 162


RESULT  29
```

```
EAT3_MOUSE
ID   EAT3_MOUSE     STANDARD;      PRT;   523 AA.
AC   P51906; O35869; O35875;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENT
DE   GLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY AMINO-ACID CARRIER 1).
GN   SLC1A1 OR EAAT3 OR EAAC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCL:ICR; TISSUE=Cerebellum;
RX   MEDLINE=95284091; PubMed=7766664;
RA   Mukainaka Y., Tanaka K., Hagiwara T., Wada K.;
RT   "Molecular cloning of two glutamate transporter subtypes from mouse
RT   brain.";
RL   Biochim. Biophys. Acta 1244:233-237(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Peng J.-B., Guo L.-H.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC       ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC       RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC       AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG, KIDNEY AND SKELETAL
CC       MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC       (SDF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D43797; BAA07855.1; -.
DR   EMBL; U73521; AAB80694.1; -.
DR   EMBL; U75217; AAB71742.1; -.
DR   MGD; MGI:105083; Slc1a1.
DR   InterPro; IPR001991; -.
DR   Pfam; PF00375; SDF; 1.
DR   PRINTS; PR00173; EDTRNSPORT.
DR   PROSITE; PS00713; NA_DICARBOXYL_SYMP_1; 1.
DR   PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
KW   Transport; Transmembrane; Glycoprotein; Symport; Multigene family.
FT   DOMAIN       1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19    38       POTENTIAL.
FT   TRANSMEM     62    82       POTENTIAL.
FT   TRANSMEM     94    114      POTENTIAL.
FT   DOMAIN       115   208      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     209   228      POTENTIAL.
FT   TRANSMEM     253   272      POTENTIAL.
FT   TRANSMEM     289   308      POTENTIAL.
FT   TRANSMEM     315   334      POTENTIAL.
FT   TRANSMEM     361   380      POTENTIAL.
FT   TRANSMEM     391   410      POTENTIAL.
FT   TRANSMEM     419   438      POTENTIAL.
FT   CARBOHYD     128   128      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     178   178      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     194   194      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     121   121      T -> A (IN AAB71742).
FT   CONFLICT     139   139      I -> M (IN REF. 2).
SQ   SEQUENCE   523 AA;  56675 MW;  F20AD35447F398FB CRC64;
```

```
  Query Match              1.6%;  Score 7;  DB 1;  Length 523;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      219 TVAAVVL 225
            |||||||
Db       24 TVAAVVL 30


RESULT  30
EAT3_RAT
ID   EAT3_RAT       STANDARD;      PRT;   523 AA.
AC   P51907; Q63727;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENT
DE   GLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY AMINO-ACID CARRIER 1).
GN   SLC1A1 OR EAAT3 OR EAAC1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RA   Bjoras M., Gjesdal O., Erickson J.D., Torp R., Levy L.M.,
RA   Ottersen O.P., Degre M., Storm-Mathisen J., Seeberg E.,
RA   Danbolt N.C.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=96318324; PubMed=8747153;
RA   Kanai Y., Bhide P.G., Difiglia M., Hediger M.A.;
RT   "Neuronal high-affinity glutamate transport in the rat central
RT   nervous system.";
RL   NeuroReport 6:2357-2362(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Brain;
RA   Kiryu S., Yao G.L., Morita N., Kato H., Kiyama H.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 156-368 FROM N.A.
RX   MEDLINE=96346642; PubMed=8738164;
RA   Palos T.P., Ramachandran B., Boado R., Howard B.D.;
RT   "Rat C6 and human astrocytic tumor cells express a neuronal type of
RT   glutamate transporter.";
RL   Brain Res. Mol. Brain Res. 37:297-303(1996).
CC   -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC       ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC       RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC       AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC       (SDF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X94255; CAA63937.1; -.
DR   EMBL; U39555; AAB09773.1; -.
DR   EMBL; D63772; BAA09849.1; -.
DR   EMBL; L35558; AAB51161.1; -.
DR   InterPro; IPR001991; -.
DR   Pfam; PF00375; SDF; 1.
```

```
DR   PRINTS; PR00173; EDTRNSPORT.
DR   PROSITE; PS00713; NA_DICARBOXYL_SYMP_1; 1.
DR   PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
KW   Transport; Transmembrane; Glycoprotein; Symport; Multigene family.
FT   DOMAIN        1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19     38       POTENTIAL.
FT   TRANSMEM     62     82       POTENTIAL.
FT   TRANSMEM     94    114       POTENTIAL.
FT   DOMAIN      115    208       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    209    228       POTENTIAL.
FT   TRANSMEM    253    272       POTENTIAL.
FT   TRANSMEM    289    308       POTENTIAL.
FT   TRANSMEM    315    334       POTENTIAL.
FT   TRANSMEM    361    380       POTENTIAL.
FT   TRANSMEM    391    410       POTENTIAL.
FT   TRANSMEM    419    438       POTENTIAL.
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     68     68       L -> M (IN REF. 3).
FT   CONFLICT     93     93       R -> C (IN REF. 3).
FT   CONFLICT    149    151       MFP -> ILG (IN REF. 2).
FT   CONFLICT    156    156       Q -> E (IN REF. 4).
FT   CONFLICT    203    203       G -> C (IN REF. 4).
FT   CONFLICT    218    218       C -> F (IN REF. 4).
FT   CONFLICT    309    309       R -> T (IN REF. 4).
FT   CONFLICT    383    383       L -> V (IN REF. 3).
SQ   SEQUENCE   523 AA;  56771 MW;  D6783244E37640E4 CRC64;

  Query Match             1.6%;  Score 7;  DB 1;  Length 523;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     219 TVAAVVL 225
           |||||||
Db      24 TVAAVVL 30


RESULT  31
EAT3_HUMAN
ID   EAT3_HUMAN     STANDARD;      PRT;   524 AA.
AC   P43005;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENT
DE   GLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY AMINO-ACID CARRIER 1).
GN   SLC1A1 OR EAAT3 OR EAAC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=94365697; PubMed=7521911;
RA   Arriza J.L., Fairman W.A., Wendy A., Wadiche J.I., Murdoch G.H.,
RA   Kavanaugh M.P., Amara S.G.;
RT   "Functional comparisons of three glutamate transporter subtypes
RT   cloned from human motor cortex.";
RL   J. Neurosci. 14:5559-5569(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95162720; PubMed=7859077;
RA   Shashidharan P., Huntley G.W., Meyer T., Morrison J.H.,
RA   Plaitakis A.;
RT   "Neuron-specific human glutamate transporter: molecular cloning,
RT   characterization and expression in human brain.";
RL   Brain Res. 662:245-250(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94327639; PubMed=7914198;
RA   Kanai Y., Stelzner M., Nussberger S., Khawaja S., Hebert S.C.,
RA   Smith C.P., Hediger M.A.;
RT   "The neuronal and epithelial human high affinity glutamate
RT   transporter. Insights into structure and mechanism of transport.";
RL   J. Biol. Chem. 269:20599-20606(1994).
CC   -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC       ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC       RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC       AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC       (SDF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U03506; AAA50430.1; -.
DR   EMBL; U08989; AAA68628.1; -.
DR   EMBL; U06469; AAA53215.1; -.
DR   MIM; 133550; -.
DR   InterPro; IPR001991; -.
DR   Pfam; PF00375; SDF; 1.
DR   PRINTS; PR00173; EDTRNSPORT.
DR   PROSITE; PS00713; NA_DICARBOXYL_SYMP_1; 1.
DR   PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
KW   Transport; Transmembrane; Glycoprotein; Symport; Multigene family.
FT   DOMAIN        1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19     38       POTENTIAL.
FT   TRANSMEM     62     82       POTENTIAL.
FT   TRANSMEM     94    114       POTENTIAL.
FT   DOMAIN      115    209       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    210    229       POTENTIAL.
FT   TRANSMEM    254    273       POTENTIAL.
FT   TRANSMEM    290    309       POTENTIAL.
FT   TRANSMEM    316    335       POTENTIAL.
FT   TRANSMEM    362    381       POTENTIAL.
FT   TRANSMEM    392    411       POTENTIAL.
FT   TRANSMEM    420    439       POTENTIAL.
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     10     10       E -> PS (IN REF. 1).
FT   CONFLICT     94     94       A -> V (IN REF. 2).
FT   CONFLICT     96     96       V -> L (IN REF. 3).
FT   CONFLICT    170    170       K -> N (IN REF. 3).
FT   CONFLICT    172    172       P -> A (IN REF. 3).
FT   CONFLICT    260    260       L -> H (IN REF. 3).
FT   CONFLICT    280    280       R -> A (IN REF. 2).
SQ   SEQUENCE   524 AA;  57100 MW;  FC6244ADC9EA228F CRC64;

  Query Match             1.6%;  Score 7;  DB 1;  Length 524;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     219 TVAAVVL 225
           |||||||
Db      24 TVAAVVL 30


RESULT  32
CH61_MYCTU
ID   CH61_MYCTU     STANDARD;      PRT;   539 AA.
AC   Q59573; Q59581;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
```

```
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   60 KDA CHAPERONIN 1 (PROTEIN CPN60-1) (GROEL PROTEIN 1).
GN   GROEL1 OR GROEL-1 OR RV3417C OR MTCY78.12.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ERDMANN;
RX   MEDLINE=93219332; PubMed=7681982;
RA   Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;
RT   "Mycobacterium tuberculosis expresses two chaperonin-60 homologs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612(1993).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z77165; CAB01006.1; -.
DR   EMBL; X60350; CAA42909.1; -.
DR   HSSP; P06139; 1GRL.
DR   TubercuList; Rv3417c; -.
DR   InterPro; IPR001844; -.
DR   InterPro; IPR002423; -.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Multigene family; Antigen; Cell wall.
FT   VARIANT     467    467       N -> K (IN STRAIN ERDMANN).
SQ   SEQUENCE   539 AA;  55877 MW;  3E0B93164C091B63 CRC64;

  Query Match             1.6%;  Score 7;  DB 1;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     409 LGERLAK 415
           |||||||
Db     363 LGERLAK 369


RESULT  33
GSHC_TOBAC
ID   GSHC_TOBAC     STANDARD;      PRT;   557 AA.
AC   P80461;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLUTATHIONE REDUCTASE, CHLOROPLAST PRECURSOR (EC 1.6.4.2) (GR)
DE   (GRASE) (FRAGMENT).
GN   GOR.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96081527; PubMed=8547821;
RA   Creissen G.P., Mullineaux P.M.;
RT   "Cloning and characterisation of glutathione reductase cDNAs and
RT   identification of two genes encoding the tobacco enzyme.";
RL   Planta 197:422-425(1995).
RN   [2]
RP   SEQUENCE OF 67-74.
RC   TISSUE=Leaf;
RA   Willows R.D., Kannangara G.C., Svendsen I.;
RL   Submitted (JUN-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MAINTAIN HIGH LEVELS OF REDUCED GLUTATHIONE IN THE
CC       CHLOROPLAST.
CC   -!- CATALYTIC ACTIVITY: NADPH + OXIDIZED GLUTATHIONE = NADP(+) +
CC       2 GLUTATHIONE.
CC   -!- COFACTOR: FAD.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC   -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC       OXIDOREDUCTASES CLASS-I.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76293; CAA53925.1; -.
DR   HSSP; P00390; 1ALG.
DR   InterPro; IPR001100; -.
DR   Pfam; PF00070; pyr_redox; 1.
DR   PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW   Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP;
KW   Chloroplast; Transit peptide.
FT   NON_TER       1      1
FT   TRANSIT      <1     66       CHLOROPLAST.
FT   CHAIN        67    557       GLUTATHIONE REDUCTASE.
FT   NP_BIND      81    111       FAD (ADP PART) (BY SIMILARITY).
FT   DISULFID    127    132       REDOX-ACTIVE (BY SIMILARITY).
FT   BINDING     283    283       NADP (BY SIMILARITY).
FT   BINDING     289    289       NADP (BY SIMILARITY).
FT   NP_BIND     378    388       FAD (FLAVIN PART) (BY SIMILARITY).
FT   ACT_SITE    521    521       BY SIMILARITY.
SQ   SEQUENCE   557 AA;  60035 MW;  79813D1C6AA1D784 CRC64;

  Query Match             1.6%;  Score 7;  DB 1;  Length 557;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      73 RSSPSEG 79
           |||||||
Db     539 RSSPSEG 545


RESULT  34
ROL_HUMAN
ID   ROL_HUMAN      STANDARD;      PRT;   558 AA.
AC   P14866;
DT   01-APR-1990 (Rel. 14, Created)
```

```
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L (HNRNP L).
GN   HNRPL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90078296; PubMed=2687284;
RA   Pinol-Roma S., Swanson M.S., Gall J.G., Dreyfuss G.;
RT   "A novel heterogeneous nuclear RNP protein with a unique distribution
RT   on nascent transcripts.";
RL   J. Cell Biol. 109:2575-2587(1989).
RN   [2]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Keratinocytes;
RX   MEDLINE=93162043; PubMed=1286667;
RA   Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA   Vandekerckhove J.;
RT   "Microsequences of 145 proteins recorded in the two-dimensional gel
RT   protein database of normal human epidermal keratinocytes.";
RL   Electrophoresis 13:960-969(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE HETEROGENOUS NUCLEAR
CC       RIBONUCLEOPROTEIN (HNRNP) COMPLEXES WHICH PROVIDE THE SUBSTRATE
CC       FOR THE PROCESSING EVENTS THAT PRE-MRNAS UNDERGO BEFORE BECOMING
CC       FUNCTIONAL, TRANSLATABLE MRNAS IN THE CYTOPLASM. L IS ASSOCIATED
CC       WITH MOST NASCENT TRANSCRIPTS INCLUDING THOSE OF THE LANDMARK
CC       GIANT LOOPS OF AMPHIBIAN LAMPBRUSH CHROMOSOMES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOPLASM.
CC   -!- PTM: SEVERAL ISOELECTRIC FORMS OF THE L PROTEIN ARE PROBABLY THE
CC       RESULTS OF POSTTRANSLATIONAL MODIFICATIONS.
CC   -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16135; CAA34261.1; -.
DR   PIR; A33616; A33616.
DR   HSSP; P19339; 2SXL.
DR   SWISS-2DPAGE; P14866; HUMAN.
DR   Aarhus/Ghent-2DPAGE; 1505; IEF.
DR   Aarhus/Ghent-2DPAGE; 4602; NEPHGE.
DR   MIM; 603083; -.
DR   MIM; 164021; -.
DR   InterPro; IPR000504; -.
DR   Pfam; PF00076; rrm; 3.
DR   PROSITE; PS50102; RRM; 3.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
FT   DOMAIN        8     58       GLY-RICH.
FT   DOMAIN       71    145       RNA-BINDING (RRM) 1.
FT   DOMAIN      162    239       RNA-BINDING (RRM) 2.
FT   DOMAIN      351    425       RNA-BINDING (RRM) 3.
FT   DOMAIN      335    342       POLY-PRO.
SQ   SEQUENCE   558 AA;  60187 MW;  395E5A04B14C848D CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     404 DALETLG 410
           |||||||
Db     523 DALETLG 529
```

```
RESULT  35
CCMF_PSEFL
ID   CCMF_PSEFL     STANDARD;      PRT;   660 AA.
AC   P52225;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK.
GN   CYCK.
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=294;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=09906;
RA   Yang C.H., Azad H.R., Cooksey D.A.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC       POSSIBLE SUBUNIT OF A HEME LYASE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U44827; AAC44226.1; -.
DR   InterPro; IPR002541; -.
DR   Pfam; PF01578; CytC_asm; 1.
KW   Cytochrome c-type biogenesis; Transmembrane; Inner membrane.
FT   TRANSMEM     15     35       POTENTIAL.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM     99    119       POTENTIAL.
FT   TRANSMEM    126    146       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
FT   TRANSMEM    216    236       POTENTIAL.
FT   TRANSMEM    250    270       POTENTIAL.
FT   TRANSMEM    279    299       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    360    380       POTENTIAL.
FT   TRANSMEM    395    415       POTENTIAL.
FT   TRANSMEM    426    446       POTENTIAL.
FT   TRANSMEM    450    470       POTENTIAL.
FT   TRANSMEM    493    513       POTENTIAL.
FT   TRANSMEM    618    638       POTENTIAL.
SQ   SEQUENCE   660 AA;  72172 MW;  3C6F4B578C56436B CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 660;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 LVLVVAA 44
           |||||||
Db     361 LVLVVAA 367


RESULT  36
P115_CHICK
ID   P115_CHICK     STANDARD;      PRT;   762 AA.
AC   Q98917;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MELANOSOMAL MATRIX PROTEIN 115 KDA PROTEIN PRECURSOR.
GN   MMP115.
OS   Gallus gallus (Chicken).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Retinal pigment epithelium;
RX   MEDLINE=92020667; PubMed=1924173;
RA   Mochii M., Agata K., Eguchi G.;
RT   "Complete sequence and expression of a cDNA encoding a chicken
RT   115-kDa melanosomal matrix protein.";
RL   Pigment Cell Res. 4:41-47(1991).
RN   [2]
RP   CHARACTERIZATION.
RC   STRAIN=WHITE LEGHORN; TISSUE=Retinal pigment epithelium;
RX   MEDLINE=88311098; PubMed=3409326;
RA   Mochii M., Agata K., Kobayashi H., Yamamoto T.S., Eguchi G.;
RT   "Expression of gene coding for a melanosomal matrix protein
RT   transcriptionally regulated in the transdifferentiation of chick
RT   embryo pigmented epithelial cells.";
RL   Cell Differ. 24:67-74(1988).
CC   -!- FUNCTION: MIGHT BE REQUIRED FOR POLYMERIZATION OF MELANIN ONTO THE
CC       CORE STRUCTURE OF MELANOSOMES WITH ENZYMIC FUNCTION OF TYROSINASE.
CC   -!- SUBCELLULAR LOCATION: ON THE FIBROUS MATRIX STRUCTURE OF THE
CC       PREMELANOSOME.
CC   -!- TISSUE SPECIFICITY: SPECIFIC TO PIGMENTED EPITHELIAL CELLS AND
CC       MELANOCYTES. NOT EXPRESSED IN LENS, NEURAL RETINA, BRAIN, HEART,
CC       GIZZARD OR LIVER.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSEDED DURING THE REDIFFERENTIATION OF
CC       PIGMENTED EPITHELIAL CELLS (PEC).
CC   -!- PTM: GLYCOSYLATED.
CC   -!- SIMILARITY: BELONGS TO THE PMEL-17/NMB FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PKD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D88348; BAA13589.1; -.
DR   InterPro; IPR000601; -.
DR   Pfam; PF00801; PKD; 1.
DR   PROSITE; PS50093; PKD; 1.
KW   Signal; Glycoprotein; Repeat.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    762       MELANOSOMAL MATRIX PROTEIN 115 KDA
FT                                PROTEIN.
FT   DOMAIN      223    323       PKD.
FT   DOMAIN      441    532       4 X 20-24 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      441    464       1.
FT   REPEAT      465    488       2.
FT   REPEAT      489    508       3.
FT   REPEAT      509    532       4.
FT   CARBOHYD    111    111       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    346    346       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   762 AA;  77356 MW;  172C8DDB4FDCE7C6 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 762;
  Best Local Similarity   100.0%;  Pred. No. 94;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     294 PAEPTGV 300
           |||||||
Db     368 PAEPTGV 374
```

```
RESULT  37
SC18_CANAL
ID   SC18_CANAL     STANDARD;      PRT;   794 AA.
AC   P34732;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   VESICULAR-FUSION PROTEIN SEC18.
GN   SEC18.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 26555;
RX   MEDLINE=94025979; PubMed=8212895;
RA   Nieto A., Sanz P., Sentandreu R., del Castilo Agudo L.;
RT   "Cloning and characterization of the SEC18 gene from Candida
RT   albicans.";
RL   Yeast 9:875-887(1993).
CC   -!- FUNCTION: REQUIRED FOR VESICLE-MEDIATED TRANSPORT. CATALYZES THE
CC       FUSION OF TRANSPORT VESICLES WITHIN THE GOLGI CISTERNAE. IS ALSO
CC       REQUIRED FOR TRANSPORT FROM THE ENDOPLASMIC RETICLUM TO THE GOLGI
CC       STACK. SEEM TO FUNCTION AS A FUSION PROTEIN REQUIRED FOR THE
CC       DELIVERY OF CARGO PROTEINS TO ALL COMPARTMENTS OF THE GOLGI STACK
CC       INDEPENDENT OF VESICLE ORIGIN. BINDS TO SEC17.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66467; CAA47077.1; -.
DR   PIR; S37606; S37606.
DR   InterPro; IPR001939; -.
DR   Pfam; PF00004; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Transport; Protein transport; Endoplasmic reticulum; Golgi stack;
KW   ATP-binding; Repeat.
FT   NP_BIND     317    324       ATP (POTENTIAL).
FT   NP_BIND     600    607       ATP (POTENTIAL).
SQ   SEQUENCE   794 AA;  88925 MW;  35A22E62C196F67D CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 794;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     278 ILQPTQV 284
           |||||||
Db     173 ILQPTQV 179


RESULT  38
RRPO_SBMV
ID   RRPO_SBMV      STANDARD;      PRT;   956 AA.
AC   P21405;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PROBABLE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48).
OS   Southern bean mosaic virus (SBMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Sobemovirus.
OX   NCBI_TaxID=12139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COWPEA;
```

```
RX   MEDLINE=88044510; PubMed=2823471;
RA   Wu S., Rinehart C.A., Kaesberg P.;
RT   "Sequence and organization of southern bean mosaic virus genomic
RT   RNA.";
RL   Virology 161:73-80(1987).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23021; AAA46565.1; -.
DR   PIR; B33739; RRBWSC.
DR   InterPro; IPR001795; -.
DR   Pfam; PF02123; Luteo_ORF3; 1.
DR   PRINTS; PR00914; LVIRUSRNAPOL.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   956 AA;  104844 MW;  ABB981C9B76082D8 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 956;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     199 TSSPGTP 205
           |||||||
Db     516 TSSPGTP 522


RESULT  39
RPB2_DROME
ID   RPB2_DROME     STANDARD;      PRT;  1176 AA.
AC   P08266; Q95027; Q04155; Q9VFM7;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE II 140 KDA POLYPEPTIDE (EC 2.7.7.6)
DE   (RNA POLYMERASE II SUBUNIT 2).
GN   RPII140 OR CG3180.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE OF 54-1176 FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=88011299; PubMed=3116266;
RA   Falkenburg D., Dworniczak B., Faust D.M., Bautz E.K.F.;
RT   "RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit
RT   to the beta subunit of Escherichia coli RNA polymerase.";
RL   J. Mol. Biol. 195:929-937(1987).
RN   [2]
RP   SEQUENCE OF 1-69 FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=91276237; PubMed=1905256;
RA   Sitzler S., Oldenburg I., Peterson G., Bautz E.K.F.;
RT   "Analysis of the promoter region of the housekeeping gene DmRP140 by
RT   sequence comparison of Drosophila melanogaster and Drosophila
RT   virilis.";
RL   Gene 100:155-162(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14
CC       DIFFERENT POLYPEPTIDES. RNA POLYMERASE II CONSISTS OF 10 DIFFERENT
CC       SUBUNITS. THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA
CC       POLYMERASE II.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05709; CAA29180.2; -.
DR   EMBL; M62972; AAA28476.1; -.
DR   EMBL; AE003703; AAF55024.1; -.
DR   PIR; A27826; A27826.
DR   FlyBase; FBgn0003276; RpII140.
DR   InterPro; IPR001572; -.
DR   Pfam; PF00562; RNA_pol_B; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW   Zinc-finger; Metal-binding; Nuclear protein.
FT   ZN_FING    1121   1142       C4-TYPE (POTENTIAL).
FT   CONFLICT     72     72       A -> R (IN REF. 1).
FT   CONFLICT    666    667       ID -> MY (IN REF. 1).
SQ   SEQUENCE   1176 AA;  134042 MW;  224821B335BED7F0 CRC64;
```

```
  Query Match             1.6%;  Score 7;  DB 1;  Length 1176;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      25 ARGARPG 31
           |||||||
Db     324 ARGARPG 330


RESULT  40
POLS_SINDO
ID   POLS_SINDO     STANDARD;      PRT;  1245 AA.
AC   P27285; Q00349;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   STRUCTURAL POLYPROTEIN (P130) [CONTAINS: COAT PROTEIN C (EC 3.4.21.-);
DE   SPIKE GLYCOPROTEINS E3, E2 AND E1; 6 KDA PEPTIDE].
OS   Sindbis virus (subtype Ockelbo / strain Edsbyn 82-5).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=31699;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220725; PubMed=1673813;
RA   Shirako Y., Niklasson B., Dalrymple J.M., Strauss E.G., Strauss J.H.;
RT   "Structure of the Ockelbo virus genome and its relationship to other
RT   Sindbis viruses.";
RL   Virology 182:753-764(1991).
CC   -!- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC       SEQUENCE FOR THE SPIKE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC       HEMAGGLUTININ.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69205; AAA96973.1; -.
DR   EMBL; M69207; AAA73066.1; -.
DR   PIR; B39991; VHWV82.
DR   HSSP; P03316; 1KXF.
DR   MEROPS; S03.001; -.
DR   InterPro; IPR000930; -.
DR   InterPro; IPR000936; -.
DR   InterPro; IPR001836; -.
DR   InterPro; IPR002533; -.
DR   InterPro; IPR002548; -.
DR   Pfam; PF01589; Alpha_E1_glycop; 1.
DR   Pfam; PF00943; Alpha_E2_glycop; 1.
DR   Pfam; PF01563; Alpha_E3_glycop; 1.
DR   Pfam; PF00944; Alpha_core; 1.
DR   PRINTS; PR00798; TOGAVIRIN.
KW   Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
KW   Serine protease.
FT   CHAIN         1    264       COAT PROTEIN C (CAPSID PROTEIN C).
FT   CHAIN       265    328       SPIKE GLYCOPROTEIN E3.
FT   CHAIN       329    751       SPIKE GLYCOPROTEIN E2.
FT   CHAIN       752    806       6 KDA PEPTIDE.
FT   CHAIN       807   1245       SPIKE GLYCOPROTEIN E1.
FT   ACT_SITE    141    141       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    147    147       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    223    223       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   TRANSMEM    696    712       POTENTIAL.
FT   TRANSMEM    728    746       POTENTIAL.
FT   TRANSMEM    768    784       POTENTIAL.
FT   TRANSMEM    786    802       POTENTIAL.
FT   TRANSMEM   1216   1234       POTENTIAL.
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    524    524       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    646    646       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    945    945       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1051   1051       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1245 AA;  136649 MW;  967EF00E675F84EF CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     216 IGVTVAA 222
           |||||||
Db     708 IGVTVAA 714


RESULT  41
BIMB_EMENI
ID   BIMB_EMENI     STANDARD;      PRT;  2067 AA.
AC   P33144;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   CELL DIVISION-ASSOCIATED PROTEIN BIMB.
GN   BIMB.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A773;
RX   MEDLINE=92348436; PubMed=1639810;
RA   May G.S., McGoldrick C.A., Holt C.L., Denison S.H.;
RT   "The bimB3 mutation of Aspergillus nidulans uncouples DNA replication
RT   from the completion of mitosis.";
RL   J. Biol. Chem. 267:15737-15743(1992).
CC   -!- FUNCTION: REQUIRED FOR NUCLEAR DIVISION. COULD FUNCTION IN THE
CC       MITOTIC SPINDLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: TO YEAST ESP1 AND S.POMBE CUT1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M83232; AAA33297.1; ALT_TERM.
DR   PIR; A42854; A42854.
DR   MEROPS; C50.001; -.
KW   Nuclear protein; Mitosis.
SQ   SEQUENCE   2067 AA;  227925 MW;  E0655D939EC148DB CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 2067;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     306 GESEHLL 312
           |||||||
Db    1417 GESEHLL 1423


RESULT  42
RRPL_NDVB
ID   RRPL_NDVB      STANDARD;      PRT;  2204 AA.
AC   P11205;
DT   01-JUL-1989 (Rel. 11, Created)
```

```
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Newcastle disease virus (strain Beaudette C/45) (NDV).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11178;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87230982; PubMed=3035486;
RA   Yusoff K., Millar N.S., Chambers P., Emmerson P.T.;
RT   "Nucleotide sequence analysis of the L gene of Newcastle disease
RT   virus: homologies with Sendai and vesicular stomatitis viruses.";
RL   Nucleic Acids Res. 15:3961-3976(1987).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05399; CAA28985.1; -.
DR   PIR; A26747; RRNZNV.
DR   InterPro; IPR001016; -.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2204 AA;  248822 MW;  C67B8674D904802C CRC64;
```

```
  Query Match             1.6%;  Score 7;  DB 1;  Length 2204;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     116 CTRCDSG 122
           |||||||
Db    1157 CTRCDSG 1163
```

```
RESULT  43
PGG2_RAT
ID   PGG2_RAT       STANDARD;      PRT;  2326 AA.
AC   Q00657;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CHONDROITIN SULFATE PROTEOGLYCAN NG2 PRECURSOR (HSN TUMOUR-SPECIFIC
DE   ANTIGEN).
GN   NG2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 30-47; 1011-1016 AND 1466-1477.
RC   TISSUE=Neuron;
RX   MEDLINE=91302479; PubMed=1906475;
RA   Nishiyama A., Dahlin K.J., Prince J.T., Johnstone S.R., Stallcup W.B.;
RT   "The primary structure of NG2, a novel membrane-spanning
RT   proteoglycan.";
RL   J. Cell Biol. 114:359-371(1991).
RN   [2]
RP   REVISIONS TO 2047-2096.
RA   Nishiyama A., Dahlin K.J., Prince J.T., Johnstone S.R., Stallcup W.B.;
RL   J. Cell Biol. 145:1115-1115(1999).
RN   [3]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 556-573 AND 657-667.
RX   MEDLINE=94357704; PubMed=8077056;
RA   Leger O., Johnson-Leger C., Jackson E., Coles B., Dean C.;
RT   "The chondroitin sulfate proteoglycan NG2 is a tumour-specific antigen
RT   on the chemically induced rat chondrosarcoma HSN.";
RL   Int. J. Cancer 58:700-705(1994).
CC   -!- FUNCTION: NG2 IS A LARGE CHONDROITIN SULFATE PROTEOGLYCAN OF
CC       400-800 KDA WITH A CORE PROTEIN SIZE OF 300 KDA.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: NEURAL CELLS AND ALSO EXTRANEURAL TISSUES,
CC       ESPECIALLY IN THE DEVELOPING MESENCHYME.
CC   -!- DEVELOPMENTAL STAGE: THE LEVEL OF EXPRESSION IS HIGHEST ON
CC       IMMATURE, PROLIFERATING CELLS AND DECREASES AS THESE CELLS
CC       DIFFERENTIATE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56541; CAA39884.2; -.
DR   PIR; S16025; S16025.
DR   InterPro; IPR001791; -.
DR   Pfam; PF00054; laminin_G; 1.
KW   Glycoprotein; Transmembrane; Proteoglycan; Signal; Repeat.
FT   SIGNAL        1     29
FT   CHAIN        30   2326       CHONDROITIN SULFATE PROTEOGLYCAN NG2.
FT   DOMAIN       30   2225       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2226   2250       POTENTIAL.
FT   DOMAIN     2251   2326       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       30    640       GLOBULAR OR COMPACT CONFIGURATION
FT                                STABILIZED BY DISULFIDE BONDS.
FT   DOMAIN      384    391       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      640   1590       GLY/SER-RICH (GLYCOSAMINOGLYCAN
FT                                ATTACHMENT DOMAIN) (POTENTIAL).
FT   DOMAIN     1591   2225       CYSTEINE-CONTAINING.
FT   DOMAIN      506   1932       4 X REPEATS, A-TYPE.
FT   REPEAT      506    650       A1.
FT   REPEAT      975   1118       A2.
FT   REPEAT     1435   1571       A3.
FT   REPEAT     1789   1932       A4.
FT   DOMAIN      506   1932       2 X REPEATS, B-TYPE.
FT   REPEAT     1535   1564       B1.
FT   REPEAT     2001   2030       B2.
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    349    349       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    428    428       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    773    773       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    999    999       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD   1135   1135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1206   1206       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1368   1368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1453   1453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1649   1649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1913   1913       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2020   2020       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2038   2038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2044   2044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2079   2079       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1391   1392       LG -> FP (IN REF. 3).
SQ   SEQUENCE   2326 AA;  251907 MW;  FA3E3BCEDB7E5EF5 CRC64;
```

```
  Query Match             1.6%;  Score 7;  DB 1;  Length 2326;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     241 LPYLKGI 247
```

```
          |||||||
Db    157 LPYLKGI 163


RESULT  44
POLG_TEV
ID   POLG_TEV       STANDARD;      PRT;  3054 AA.
AC   P04517; Q88500; Q88501; Q88502; Q88504; Q88505; Q88506; Q89773;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE   COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
DE   1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
DE   GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE   (EC 3.4.22.-) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
DE   PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE   COAT PROTEIN (CP)].
OS   Tobacco etch virus (TEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC   Potyvirus.
OX   NCBI_TaxID=12227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Allison R., Johnston R.E., Dougherty W.G.;
RT   "The nucleotide sequence of the coding region of tobacco etch virus
RT   genomic RNA: evidence for the synthesis of a single polyprotein.";
RL   Virology 154:9-20(1986).
RN   [2]
RP   SEQUENCE OF 2344-3054 FROM N.A.
RA   Allison R.F., Sorenson J.C., Kelly M.E., Armstrong F.B.,
RA   Dougherty W.G.;
RT   "Sequence determination of the capsid protein gene and flanking
RT   regions of tobacco etch virus: evidence for synthesis and processing
RT   of a polyprotein in potyvirus genome expression.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:3969-3972(1985).
RN   [3]
RP   IDENTIFICATION OF PROTEASES.
RX   MEDLINE=89251560; PubMed=2656254;
RA   Carrigton J.C., Cary S.M., Parks T.D., Dougherty W.G.;
RT   "A second proteinase encoded by a plant potyvirus genome.";
RL   EMBO J. 8:365-370(1989).
CC   -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC       TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC   -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC       MAY BE INVOLVED IN REPLICATION.
CC   -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC   -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC   -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC       POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC       PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC       INDIVIDUAL PROTEINS.
CC   -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC   -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC   -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15239; AAA47910.1; -.
DR   EMBL; M11458; AAA47909.1; -.
DR   EMBL; M11216; AAA47908.1; ALT_SEQ.
DR   PIR; A04207; GNBVEV.
DR   MEROPS; C04.001; -.
DR   MEROPS; C06.001; -.
DR   MEROPS; S30.001; -.
DR   InterPro; IPR001205; -.
DR   InterPro; IPR001410; -.
DR   InterPro; IPR001456; -.
DR   InterPro; IPR001592; -.
DR   InterPro; IPR001730; -.
DR   InterPro; IPR002540; -.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00863; Peptidase_C4; 1.
DR   Pfam; PF00851; Peptidase_C6; 1.
DR   Pfam; PF01577; Poty_P1; 1.
DR   Pfam; PF00767; Poty_coat; 1.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   PRINTS; PR00966; NIAPOTYPTASE.
KW   Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW   Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW   ATP-binding.
FT   CHAIN         1      ?       N-TERMINAL PROTEIN.
FT   CHAIN         ?    763       HELPER COMPONENT PROTEINASE.
FT   CHAIN       764      ?       PROTEIN P3.
FT   CHAIN         ?   1163       6 KDA PROTEIN 1.
FT   CHAIN      1164   1796       CYTOPLASMIC INCLUSION PROTEIN.
FT   CHAIN      1797   1849       6 KDA PROTEIN 2.
FT   CHAIN      1850      ?       GENOME-LINKED PROTEIN.
FT   CHAIN         ?   2279       NUCLEAR INCLUSION PROTEIN A.
FT   CHAIN      2280   2791       NUCLEAR INCLUSION PROTEIN B.
FT   CHAIN      2792   3054       COAT PROTEIN.
FT   SITE        763    764       CLEAVAGE (BY HC-PRO PROTEASE).
FT   SITE       1163   1164       CLEAVAGE (BY 49 KDA PROTEASE).
FT   SITE       1796   1797       CLEAVAGE (BY 49 KDA PROTEASE).
FT   SITE       1849   1850       CLEAVAGE (BY 49 KDA PROTEASE).
FT   SITE       2279   2280       CLEAVAGE (BY 49 KDA PROTEASE).
FT   SITE       2791   2792       CLEAVAGE (BY 49 KDA PROTEASE).
FT   BINDING    1911   1911       COVALENT LINKAGE OF VIRAL RNA (BY
FT                                SIMILARITY).
FT   NP_BIND    1247   1254       ATP (POTENTIAL).
SQ   SEQUENCE   3054 AA;  346160 MW;  0AF9A3626960B5CE CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 3054;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 REARGAR 29
           |||||||
Db    1861 REARGAR 1867


RESULT  45
POLG_LMV0
ID   POLG_LMV0      STANDARD;      PRT;  3255 AA.
AC   P31999; P90263;
DT   01-JUL-1993 (Rel. 26, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE   COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
DE   1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
DE   GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE   (EC 3.4.22.-) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
DE   PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE   COAT PROTEIN (CP)].
OS   Lettuce mosaic virus (strain 0 / French isolate) (LMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC   Potyvirus.
OX   NCBI_TaxID=117131;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97239892; PubMed=9085548;
RA   Revers F., Yang S.J., Walter J., Souche S., Lot H., le Gall O.,
RA   Candresse T., Dunez J.;
RT   "Comparison of the complete nucleotide sequences of two isolates of
RT   lettuce mosaic virus differing in their biological properties.";
RL   Virus Res. 47:167-177(1997).
RN   [2]
RP   SEQUENCE OF 2776-3255 FROM N.A.
```

```
RX   MEDLINE=91158469; PubMed=2001176;
RA   Dinant S., Lot H., Albouy J., Kuziak C., Meyer M.,
RA   Astier-Manifacier S.;
RT   "Nucleotide sequence of the 3' terminal region of lettuce mosaic
RT   potyvirus RNA shows a Gln/Val dipeptide at the cleavage site between
RT   the polymerase and the coat protein.";
RL   Arch. Virol. 116:235-252(1991).
CC   -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC       TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC   -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC       MAY BE INVOLVED IN REPLICATION.
CC   -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC   -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC   -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC       POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC       PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC       INDIVIDUAL PROTEINS.
CC   -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC   -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC   -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X97704; CAA66280.1; -.
DR   EMBL; X65652; CAA46602.1; -.
DR   PIR; A43493; A43493.
DR   InterPro; IPR001205; -.
DR   InterPro; IPR001410; -.
DR   InterPro; IPR001456; -.
DR   InterPro; IPR001592; -.
DR   InterPro; IPR001730; -.
DR   InterPro; IPR002540; -.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00863; Peptidase_C4; 1.
DR   Pfam; PF00851; Peptidase_C6; 1.
DR   Pfam; PF01577; Poty_P1; 1.
DR   Pfam; PF00767; Poty_coat; 1.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   PRINTS; PR00966; NIAPOTYPTASE.
KW   Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW   Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW   ATP-binding.
FT   CHAIN         1      ?       N-TERMINAL PROTEIN.
FT   CHAIN         ?      ?       HELPER COMPONENT PROTEINASE.
FT   CHAIN         ?      ?       PROTEIN P3.
FT   CHAIN         ?      ?       6 KDA PROTEIN 1.
FT   CHAIN         ?      ?       CYTOPLASMIC INCLUSION PROTEIN.
FT   CHAIN         ?      ?       6 KDA PROTEIN 2.
FT   CHAIN         ?      ?       GENOME-LINKED PROTEIN.
FT   CHAIN         ?      ?       NUCLEAR INCLUSION PROTEIN A.
FT   CHAIN         ?   2977       NUCLEAR INCLUSION PROTEIN B.
FT   CHAIN      2978   3255       COAT PROTEIN.
FT   NP_BIND    1410   1417       ATP (POTENTIAL).
SQ   SEQUENCE   3255 AA;  367534 MW;  CD9AD5E4A0B65953 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 3255;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     244 LKGICSG 250
           |||||||
Db    1141 LKGICSG 1147


RESULT  46
POLG_LMVE
```

```
ID   POLG_LMVE      STANDARD;      PRT;  3255 AA.
AC   P89876;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE   COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
DE   1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
DE   GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE   (EC 3.4.22.-) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
DE   PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE   COAT PROTEIN (CP)].
OS   Lettuce mosaic virus (strain E) (LMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC   Potyvirus.
OX   NCBI_TaxID=117132;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97239892; PubMed=9085548;
RA   Revers F., Yang S.J., Walter J., Souche S., Lot H., le Gall O.,
RA   Candresse T., Dunez J.;
RT   "Comparison of the complete nucleotide sequences of two isolates of
RT   lettuce mosaic virus differing in their biological properties.";
RL   Virus Res. 47:167-177(1997).
CC   -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC       TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC   -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC       MAY BE INVOLVED IN REPLICATION.
CC   -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC   -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC   -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC       POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC       PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC       INDIVIDUAL PROTEINS.
CC   -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC   -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC   -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X97705; CAA66281.1; -.
DR   MEROPS; C04.001; -.
DR   MEROPS; C06.001; -.
DR   InterPro; IPR001205; -.
DR   InterPro; IPR001410; -.
DR   InterPro; IPR001456; -.
DR   InterPro; IPR001592; -.
DR   InterPro; IPR001730; -.
DR   InterPro; IPR002540; -.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00863; Peptidase_C4; 1.
DR   Pfam; PF00851; Peptidase_C6; 1.
DR   Pfam; PF01577; Poty_P1; 1.
DR   Pfam; PF00767; Poty_coat; 1.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   PRINTS; PR00966; NIAPOTYPTASE.
KW   Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW   Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW   ATP-binding.
FT   CHAIN         1      ?       N-TERMINAL PROTEIN.
FT   CHAIN         ?      ?       HELPER COMPONENT PROTEINASE.
FT   CHAIN         ?      ?       PROTEIN P3.
FT   CHAIN         ?      ?       6 KDA PROTEIN 1.
FT   CHAIN         ?      ?       CYTOPLASMIC INCLUSION PROTEIN.
FT   CHAIN         ?      ?       6 KDA PROTEIN 2.
FT   CHAIN         ?      ?       GENOME-LINKED PROTEIN.
FT   CHAIN         ?      ?       NUCLEAR INCLUSION PROTEIN A.
```

```
FT   CHAIN        ?   2977       NUCLEAR INCLUSION PROTEIN B.
FT   CHAIN     2978   3255       COAT PROTEIN.
FT   NP_BIND   1410   1417       ATP (POTENTIAL).
SQ   SEQUENCE   3255 AA;  367618 MW;  B3E8582927E01628 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 3255;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    244 LKGICSG 250
          |||||||
Db   1141 LKGICSG 1147


RESULT  47
POLG_DEN3
ID   POLG_DEN3      STANDARD;      PRT;  3390 AA.
AC   P27915;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE   PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E;
DE   NONSTRUCTURAL PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3);
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)].
OS   Dengue virus type 3.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
OX   NCBI_TaxID=11069;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90266483; PubMed=2345967;
RA   Osatomi K., Sumiyoshi H.;
RT   "Complete nucleotide sequence of dengue type 3 virus genome RNA.";
RL   Virology 176:643-647(1990).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M93130; AAA99437.1; -.
DR   PIR; A34774; GNWVD3.
DR   HSSP; P14336; 1SVB.
DR   MEROPS; S07.002; -.
DR   InterPro; IPR000069; -.
DR   InterPro; IPR000208; -.
DR   InterPro; IPR000336; -.
DR   InterPro; IPR000404; -.
DR   InterPro; IPR000487; -.
DR   InterPro; IPR000752; -.
DR   InterPro; IPR001122; -.
DR   InterPro; IPR001157; -.
DR   InterPro; IPR001528; -.
DR   InterPro; IPR001850; -.
DR   InterPro; IPR002535; -.
DR   Pfam; PF01004; Flavi_M; 1.
DR   Pfam; PF00948; Flavi_NS1; 1.
DR   Pfam; PF01005; Flavi_NS2A; 1.
DR   Pfam; PF01002; Flavi_NS2B; 1.
DR   Pfam; PF01350; Flavi_NS4A; 1.
DR   Pfam; PF01349; Flavi_NS4B; 1.
DR   Pfam; PF00972; Flavi_NS5; 1.
DR   Pfam; PF01003; Flavi_capsid; 1.
DR   Pfam; PF00869; Flavi_glycoprot; 1.
DR   Pfam; PF00949; Flavi_helicase; 1.
DR   Pfam; PF01570; Flavi_propep; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein.
FT   CHAIN         1    114       CAPSID PROTEIN C.
FT   PROPEP      115    205
FT   CHAIN       206    280       ENVELOPE GLYCOPROTEIN M.
FT   CHAIN       281    773       MAJOR ENVELOPE PROTEIN E.
FT   CHAIN       774   1184       NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN      1185   1343       NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN      1344   1473       NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN      1474   2092       HELICASE (NS3).
FT   CHAIN      2093   2378       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      2379   2490       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      2491   3390       RNA-DIRECTED RNA POLYMERASE (NS5).
FT   NP_BIND    1667   1674       ATP (POTENTIAL).
FT   SITE       1758   1761       DEAH BOX.
FT   TRANSMEM     46     67       POTENTIAL.
FT   TRANSMEM    266    280       POTENTIAL.
FT   TRANSMEM    724    746       POTENTIAL.
FT   TRANSMEM    753    771       POTENTIAL.
FT   TRANSMEM   1156   1175       5 (POTENTIAL).
FT   DISULFID    283    310       BY SIMILARITY.
FT   DISULFID    340    396       BY SIMILARITY.
FT   DISULFID    354    385       BY SIMILARITY.
FT   DISULFID    372    401       BY SIMILARITY.
FT   DISULFID    463    563       BY SIMILARITY.
FT   DISULFID    580    611       BY SIMILARITY.
FT   CARBOHYD    183    183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    433    433       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    750    750       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    903    903       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    980    980       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1132   1132       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1188   1188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1661   1661       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2300   2300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2304   2304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2456   2456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2702   2702       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2712   2712       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3390 AA;  378057 MW;  666E8F70F1E1756E CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 3390;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     43 AAVLLLV 49
          |||||||
Db   2351 AAVLLLV 2357


RESULT  48
VG38_BPMD2
ID   VG38_BPMD2     STANDARD;      PRT;    50 AA.
AC   O64229;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENE 38 PROTEIN (GP38).
GN   38.
OS   Mycobacteriophage D29.
OC   Viruses.
OX   NCBI_TaxID=28369;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98300335; PubMed=9636706;
RA   Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT   "Genome structure of mycobacteriophage D29: implications for phage
RT   evolution.";
```

```
RL   J. Mol. Biol. 279:143-164(1998).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF022214; AAC18479.1; -.
SQ   SEQUENCE   50 AA;  4851 MW;  75BCC1A1CF2EF26E CRC64;


  Query Match             1.4%;  Score 6;  DB 1;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      44 AVLLLV 49
           ||||||
Db       5 AVLLLV 10


RESULT  49
HSP1_MACRU
ID   HSP1_MACRU     STANDARD;      PRT;    59 AA.
AC   P42142;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SPERM PROTAMINE P1.
GN   PRM1.
OS   Macropus rufus (Red kangaroo) (Megaleia rufa).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX   NCBI_TaxID=9321;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Sperm;
RX   MEDLINE=95215351; PubMed=7700877;
RA   Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;
RT   "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL   Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC   -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC       SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC       SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L35447; AAA74616.1; -.
DR   InterPro; IPR000221; -.
DR   Pfam; PF00260; protamine_P1; 1.
DR   PROSITE; PS00048; PROTAMINE_P1; 1.
KW   Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW   Testis; DNA condensation; Nuclear protein.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   59 AA;  8230 MW;  78F1AE592B4B2FA2 CRC64;


  Query Match             1.4%;  Score 6;  DB 1;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     319 RSQRRR 324
           ||||||
Db      25 RSQRRR 30


RESULT  50
HSP1_PHACI
ID   HSP1_PHACI     STANDARD;      PRT;    59 AA.
AC   P79990;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   SPERM PROTAMINE P1.
GN   PRM1.
OS   Phascolarctos cinereus (Koala).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Diprotodontia; Phascolarctidae; Phascolarctos.
OX   NCBI_TaxID=38626;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC       SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC       SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U87789; AAB88911.1; -.
DR   InterPro; IPR000221; -.
DR   Pfam; PF00260; protamine_P1; 1.
DR   PROSITE; PS00048; PROTAMINE_P1; 1.
KW   Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW   Testis; DNA condensation; Nuclear protein.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   59 AA;  8271 MW;  4E1FB24E21952EA2 CRC64;


  Query Match             1.4%;  Score 6;  DB 1;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     319 RSQRRR 324
           ||||||
Db      25 RSQRRR 30


Search completed: May 23, 2001, 14:22:32
Job time: 116 sec
```

This Page Blank (uspto)